Spring 1-1-2011

Low-noise Instrumentation for Near-field Microwave Microscopy

Jonathan David Chisum
University of Colorado at Boulder, jchisum@gmail.com

Low-noise Instrumentation for Near-field Microwave Microscopy

by

Jonathan David Chisum

B.S., Seattle Pacific University, 2003

M.S., University of Colorado, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

2011

This thesis entitled:

Low-noise Instrumentation for Near-field Microwave Microscopy

written by Jonathan David Chisum

has been approved for the Department of Electrical and Computer Engineering

Zoya Popović

Dejan Filipović

Date

The final copy of this thesis has been examined by the signatories, and we

Find that both the content and the form meet acceptable presentation standards

Of scholarly work in the above mentioned discipline.

Chisum, Jonathan David (Ph.D., Electrical Engineering)

Low-noise Instrumentation for Near-field Microwave Microscopy

Thesis directed by Professor Zoya Popović

This thesis addresses circuits and systems optimized for the unique requirements of near-field microwave microscopy (NFMM). A suite of qualification measurements is conducted for the systematic characterization of the NFMM measurement system. Finally, modeling methods and quantitative analysis are performed for the interpretation of resulting measurements.

An NFMM measurement typically suffers from small signal in the presence of seemingly overwhelming white and $1/f$ noise. As such, it requires instrumentation that provides signal enhancement, noise reduction, and long-term stability. This thesis describes the design and characterization of probe circuits and probe tips which enable sensitive and high-resolution NFMM with enhanced signals. The space efficient probe circuit is designed for ease of integration and eventual MMIC implementation.

The scanning Lock-in Vector Near-field Probe (LVNP) instrument is designed for the readout of the near-field probe circuit. Selection of measurement topology for the purpose of noise reduction/mitigation is described. The LVNP is characterized with respect to noise, stability, and maximum signal sensitivity.

In summary, this thesis details the design of a complete system for near-field microwave microscopy including probe tip, probe circuit, and instrument design. Performance limitations are quantified throughout the thesis in the hope of promoting a systematic approach to NFMM instrumentation, and quantitative data analysis techniques are proposed.

Dedication

This thesis is dedicated to my wife Katie, my parents Greg and Joanne, and my grandma Betty. To Katie–your faithful support and encouragement have given me the freedom to pursue this dream and give it my best. Thank you for your adventurous spirit and courage to live the adventure together. To Mom and Dad–for always being interested in my work, and for believing in me no matter what. To Grandma Betty–for making college possible, and setting me down this path.

Personal

Acknowledgments

I owe to the current and former students in our research group a debt of gratitude for the countless hours of discussion, brain-storm, and advice. I extend a special thanks to Dr. Charles Dietlein for encouraging me to embark on this journey in the first place. To Erez Falkenstein, Mike Roeberg, and Rob Scheeler for daily discussions and for the many helpful coffee breaks and barbecues. To Drs. Evan Cullens, and Nicola Kinzie, thank you for wading through the most challenging days of graduate school with me. To Dr. John Hoversten thank you for showing me what it means to work hard and to apply my curiosity toward the completion of good work. To Drs. Luke Sankey, Milos Janković, Mike Elsbury, Negar Ehsan, and Qianli Mu, thank you for all that you taught me. To Dr. Mabel Ramírez thank you for your kind advice in my first years of graduate school.

Professional

Acknowledgments

I offer special thanks to my advisor, Dr. Zoya Popović for the five and a half years she has worked with me and enabled me to pursue my dreams. Thank you for being willing to take an inexperienced undergraduate with minimal RF background into your lab. Your instruction, your advice, and the introduction I have received into the broader world of research have been invaluable. Thank you for the conference trips, the many hours you have dedicated to reading and helping to write my papers, the many opportunities to expand my horizons (e.g., Finland), and for the truly outstanding radio astronomy field trips. Thank you for your advice throughout graduate school, publication, research, and now my professional pursuits. I look forward to working with you in the future.

I thank Dr. Erich Grossman for teaching me so many valuable lessons as I worked in his lab for the first two years of graduate school, during a time when I was arguably not a very valuable contributor. But, I feel those years were foundational to how I conduct my research today. I look forward to future collaborations.

Thank you to Drs. Pavel Kabos and T. Mitch Wallis at NIST for your help as I began my work in near-field microwave microscopy and for the helpful discussions along the way. For the time I spent in Finland, I thank Drs. Arttu Luukanen and Ville Viikari for being gracious and generous hosts.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Leonardo da Vinci was like a man who awoke too early in the darkness, while the others were all still asleep

—Sigmund Freud

Let me tell you the secret that has led me to my goal: My strength lies solely in my tenacity.

—Louis Pasteur

Contents

1.1 Overview

Scientific exploration of the physical world is replete with examples of fundamental limits discovered, then subsequently side-stepped. However, the manner in which these limits are

bypassed is such that the limit is not broken, rather techniques are found in which the limits do not apply. One such example is the well known Fourier limit [1] in which time and bandwidth are inversely related and their product is bounded below as $\Delta t \Delta f \geq 1/2$ [2]. In other words, the more a signal is concentrated in time, the more its bandwidth is dispersed. However it has been recently shown that by working around the infinite temporal support of the Fourier transform, a temporally localized super-oscillation does not incur the same penalty and can therefore slightly exceed the Fourier limit [3].

In an analogous scenario, Ernst Abbe in 1873, while formulating a theory for the resolution of an optical microscope, showed that spatial resolution is limited to one half the wavelength of the interrogating radiation, in his case propagating electromagnetic waves [4]. To the modern day this has limited visible light microscopes to 200 nm spatial resolution, commensurate with the wavelength of violet light. In the constant pursuit of extended spatial resolution in microscopy the frequency of illumination has continually increased but practical limits prevent indefinite improvement. In 1928 E. H. Synge proposed a method of high resolution microscopy that was not bounded by the Abbe barrier. This method of decoupling spatial resolution from wavelength was to be accomplished by scanning a sample through the near-field of a small probe (aperture or tip) thereby achieving sub-wavelength spatial resolution. This method today is referred to as scanning near-field microwave microscope (NFMM).

1.2 Scope of this Thesis

During the next 40 years many would attempt the measurement, or refine the theory, but not until 1965 [5] would Synge's insight be verified, and finally in 1972 the technique would achieve widespread adoption [6]. Since that time, the nascent field of near-field microwave microscopy (NFMM) has made rapid advances, first achieving millimeter resolution [5], then micrometer [7], and today hundreds of nanometers [8, 9]. NFMM has been applied to fields as diverse as biology [10], material science [11], and fundamental physics [12].

Advances in new materials, meta-materials, and micro- and nanofabricated devices and

structures in the last several decades have continued to focus on reduction of scale and vertical and lateral integration of multiple materials, and NFMM is one technique that will help to address these needs.

As fabrication capabilities have pushed beyond 45 nm to 22 nm, diagnostic tools have continued to fall behind. This has been identified as a significant need and in the last six years there have been four high profile research initiatives aimed at circuit diagnostic tools including the Defense Science Board High Performance Microchip Supply report [13], the DARPA TRUST in Integrated Circuits (TRUST-IC) program [14], the follow-on DARPA Integrity and Reliability of Integrated Circuits (IRIS) program [15], and the IARPA Circuit Analysis Tools (CAT) program [16]. In addition to circuit fabrication diagnostic tools there is a need for rapid characterization of new materials [11, 17], characterization of thin films [7, 18, 19], knowledge of semiconductor dopant profiles [20, 12], and the ability to image subsurface features of such structures [21, 22, 23]. The near-field microwave microscope (NFMM) has shown promise in helping to meet these needs [24] by providing a means for the non-destructive, localized measurement of complex material parameters (e.g., σ, ϵ, μ) of the device under test (DUT) with fine spatial resolution and sensitivity to subsurface features.

A typical NFMM system is comprised of a near-field probe tip which interacts with the DUT in a spatially confined region, a resonant probe circuit which has a strong response to small changes in the probe tip-DUT coupling, and a measurement instrument, in this work referred to as the lock-in vector near-field probe (LVNP), which provides a stimulus and measures a response of the probe circuit over a scan region of interest.

There are three NFMM measurement modalities discussed in this thesis which are useful for different purposes and often require different calibration and/or analysis methods:

(1) *Differentiation* results in either a 1D or 2D dataset (not necessarily an image) which differentiates samples or regions of the same sample. This measurement differs from imaging in that one may only perform spot measurements. A vertical scan may be employed at each test location to facilitate, e.g., more accurate material differentiation.

(2) *Analysis* includes calibration of the near-field microwave microscope, in combination with various models of the probe circuit and the physical interaction of the probe tip and the sample.

(3) *Imaging* results in a 2D image representing variation of the measured parameter over the scan area. It is a qualitative measurement of the composite interaction of all of the complex DUT parameters (σ, ϵ, μ) with the near-field probe.

We refer to a refinement of the analysis modality combined with the imaging modality as quantitative imaging. This is at the forefront of the field [25] and will be viewed as the eventual goal, but has so far only been demonstrated for special cases (e.g., analytical solutions at zero scan height [26]). Calibration and processing differs for each modality but the system and fundamental measurement are the same.

In the last two decades many groups have achieved impressive results in, e.g., spatial resolution or in sample characterization, however, innovation in probe circuit design, instrumentation development, and measurement analysis has slowed. This thesis is concerned with the design of circuits and instrumentation for NFMM, and the quantitative analysis of the resulting measurements. The goal of this work is to examine measurement system topology and then build a flexible system which facilitates integration of circuits and probe tips, and then apply appropriate standard measurements and analysis procedures. The main contributions of this thesis are:

(1) A mechanical scanning solution that supports large scan areas (mm^2) with high vertical range and simple height control.

(2) A surface mount probe circuit that supports simple integration of probe tips while maintaining high Q-factor, and is amenable to hybridization and MMIC integration.

(3) An instrument that mitigates noise in a well understood way. Here we aim to fully examine the noise environment, identify limits, and suggest methods for improvement. We also define two operating modes of the instrument with various advantages.

(4) And finally, we establish several measurement modalities and apply appropriate quantitative analysis methods to the results.

Chapter 2 discusses the measurement concept as well as the design and optimization of a microwave probe circuit. We develop an equivalent circuit model which will be useful for analysis of measurements from chapter 4. Chapter 3 discusses the design of the Lock-in Vector Near-field Probe (LVNP) as a readout system for NFMM measurements optimized for the NFMM noise environment. Chapter 4 presents measurements with the LVNP and discusses quantitative techniques for analyzing the results. Chapter 5 presents progress toward integration of the LVNP and probe circuit with monolithic microwave integrated circuits (MMICs) and presents the concept of arrays of probes for rapid and/or multi-spectral scanning. Appendix A presents a low-noise readout circuit designed for video-rate THz imaging arrays. In this appendix we develop the noise theory used throughout this work.

1.3 Concept to Realization

According to Abbe's formulation of microscope image formation [4], the resolution limit for coherent illumination is

$$|Y| = 0.82\frac{\lambda_0}{n\sin\theta}, \tag{1.1}$$

where $n\sin\theta$, the numerical aperture, is a function of the index of refraction n and the objective half-angle θ. In the limit of an infinite aperture $\theta = 90°$, and assuming practical oil-immersion techniques in which $n \leq 1.51$, the resolution limit becomes

$$|Y| = 0.82\frac{\lambda_0}{1.51\sin 90°} = 0.54\lambda_0 \approx \frac{\lambda_0}{2}. \tag{1.2}$$

Visible light spatial resolution of 200 nm has been surpassed by extreme-ultra-violet microscopes which can achieve below 38 nm spatial resolution [27], and by x-ray microscopy which can achieve 0.01 to 10 nm resolution. And yet, still the Abbe barrier is the limit to these forms of microscopy. In 1928, E. H. Synge predicted a practical limit to spatial resolution through increased illumination frequency [28]:

> It is generally accepted as an axiom of microscopy that the only way to extend resolving-power lies in the employment of light of smaller wave-lengths. Practical difficulties, however, rapidly accumulate as light of increasingly smaller wave-length is brought into service, and probably little hope is entertained of arriving at a resolution much beyond $0.1\,\mu$m, with perhaps, $0.05\,\mu$m as an extreme limit.

This prediction by E. H. Synge, in 1928 has proved to be more true than he could probably imagine. Through various heroics over the intervening 80 years, spatial resolution of light-microscopes have slowly pushed past Synge's $0.1\,\mu$m resolution barrier and now hover around his $0.05\,\mu$m practical limit. These resolutions have been achieved by employing light with ever smaller wavelength such that ultra-violet and extreme-ultra-violet radiation is employed. However, employing light sources of ever-smaller wavelength is not a universal solution. X-rays are destructive to contemporary small-scale CMOS electronics [29] and are considered carcinogenic by the United States Department of Health and Human Services [30]. Additionally, material parameters such as permittivity and permeability change over frequency so it is often desirable that a material parameter be measured at an operating frequency in the microwave regime, but without incurring the requisite $\lambda_0/2$ resolution limit.

In order to surpass the Abbe barrier spatial resolution must be decoupled from the interrogating frequency. Synge's proposal [28] did just that: by scanning a small aperture in an opaque screen over a sample, one could extend microscopic resolution into the "ultra-microscopic" region, the premise being that a small structure will necessarily enforce localized boundary conditions on the electromagnetic field thereby shaping a field with structure on the scale of the aperture. In other words, the high-spatial resolution non-propagating evanescent near-field of a small structure will replace the low-spatial resolution propagating far-field of the traditional light microscope. In Synge's words [31]:

> ...a method offers itself which lies a little outside the beaten track of microscopic work and raises various technical problems of a new kind, but which makes the attainment of a resolution of $0.01\,\mu$m, and even beyond, dependent upon a technical accomplishment which does not seem practicable at present. The idea of the method is exceedingly simple, and it has been suggested to me by a distinguished physicist that it would be of advantage to give it publicly, even though I was unable to develop it in more than an abstract way.

Today we know this distinguished physicist to be none other than Albert Einstein [31]. During

several months correspondence with Einstein, Synge attempted to convince Einstein that the Berlin Institute of Physics should undertake the experiment in an effort to overcome the new "technical accomplishments". In one letter he states

> ...the present *theoretical* limitation of the resolving power in microscopy seems to be completely removed and everything comes to depend upon *technical* perfection.

The limit of spatial resolution was no longer *fundamentally* limited, rather it was *practically* limited by what he referred to as technical perfections. All that was required for a whole world of ultra-microscopy to open up were a few key technological advances. He was even so kind as to list the four most obvious:

(1) The source of illumination must be high intensity in order for evanescent fields to appreciably interact with the sample.

(2) Motion control on the order of 10^{-6} cm laterally and 10^{-7} cm vertically.

(3) Planarity of the target: on the order of 10^{-6} cm.

(4) Construction of the small probe structure on the order of 10^{-6} cm.

In 1932 Synge suggested a method for realizing technical feat number 2, scanning. He proposed the use of piezoelectric quartz crystals for rapid and accurate positioning [32]. Today piezoelectric motion stages are almost universally employed for scanning probe microscopes (SPMs). However, there were several more achievements necessary on the road to NFMM.

In 1944, Hans Bethe developed his theory of small holes [33]. While related to, and often cited by near-field microscopy practitioners his purpose was more general. His theory indicated that the far-field of a small aperture in an opaque screen is comprised of spherical waves, but that the field intensity ($\propto E^2$) which propagates through the hole is significantly attenuated by the factor $(a/\lambda)^2$. That is, a sub-wavelength aperture will attenuate the field significantly. If an aperture is made in the side of a conducting cavity, the field around the hole would be described by higher order modes, with high spatial frequency content ("a large perturbation field exists around the hole which varies rapidly in space" [33]). Further, the fields inside will be perturbed

in such a way that the effective length of the cavity would increase and the resonant frequency would shift down accordingly. In addition, the quality factor (Q) of the cavity would go down due to increased wall losses and any radiated fields. It was Bethe's work that initially suggested the use of resonant cavities and structures in NFMM. Because of the enhanced sensitivity of a resonant probe circuit Synge's requirement for high intensity illumination was relaxed thereby addressing technical feat number 1.

As fabrication capabilities increased with every decade the ability to form small apertures and small tips became trivial addressing technical feat number 4 in Synge's 1928 list.

Target planarity and the tendency for sample topography to generate a signal much larger than the signal of interest was initially addressed by placing the sample directly on the sub-wavelength aperture. However, as NFMM systems moved from aperture-based to tip-based probes this was no longer possible. Early efforts in tip-based NFMM employed so-called "soft-contact", or scanned with such low spatial resolution (e.g., millimeter) that they did not require planarity and height control comparable to those predicted by Synge for sub-wavelength optical microscopy. However, later non-contact high resolution efforts required precise height control in a feedback loop. This was realized in two ways: optical profilometery [21], or integration with an atomic force microscope (AFM) or scanning tunneling microscope (STM) [8]. Thus, Synge's final technical feat was addressed, paving the way for rapid advancements in NFMM.

1.4 Historical Progress

After Synge's initial suggestion in 1928, and Bethe's theory in 1944, another two decades passed before the first experimental demonstration of NFMM. During this time various advances were made in the related fields of scanning electron microscopy (SEM) and scanning x-ray microscopy which tended to feed back to the early development of NFMM. In 1953 H. H. Pattee Jr. discusses recent development of a scanning x-ray microscope [34] in which he suggests that because their sample is placed immediately on top of an x-ray source radiating through a small aperture, spatial resolution is set by the extent of the aperture not the wavelength of radiation. As such, he says,

the source could be changed to a longer wavelength and the resolution of their microscope would remain the same. While this is not typically considered part of the NFMM family tree, it was an early success based on the same principle.

In 1956, O'Keefe theoretically reiterated Synge's suggestion [35] but did not contribute experimentally. O'Keefe did provide a discussion on a possible method of reading out the sub-wavelength signal. Like Synge, he envisioned a photo-detector that would be scanned around the aperture to measure the scattered light. Both were primarily concerned with sub-wavelength resolution at optical frequencies, or so-called near-field scanning optical microscopy (NSOM). The primary difference between a NFMM and a near-field scanning optical microscope (NSOM) is of course the frequency of operation, but it is often the case that the NSOM signal is read out via a scattered light detector while the NFMM signal is read out via a loaded resonator. This concept of perturbing a resonator can be traced back to Bethe's theory of small holes. Accordingly, the resonant frequency and quality factor will change not only because of the aperture (or tip), but also because of energy storage between the aperture and the sample, and losses in the sample.

During this period, experimental efforts were largely hampered by the inadequate technology so it was not uncommon to develop a low frequency analog such that scanning hardware would not limit performance. In 1956, Baez demonstrated wavelength independent resolution with acoustic waves [36]. Finally, in 1962, R. R. Soohoo performed one of the first experiments that resembled Synge's idea. It was at 5.5 GHz and achieved 25 μm resolution. Soohoo's experiment could be considered destructive though, in that it required the sample be in contact with the sub-wavelength aperture. Not until 1965 did Bryant and Gunn provide the first demonstration of a non-contact NFMM experiment in the spirit of Synge. Because of the issues associated with non-contact scanning they were only able to achieve 1 mm spatial resolution with a 450 MHz excitation [5]. The main innovation attributed to Bryant and Gunn was the concept of a guided wave through the sub-wavelength probe. As noted by Synge, Bethe, and O'Keefe, the significant attenuation associated with propagation through a near-field aperture is problematic to any NFMM implementation. Bryant and Gunn were able to measure contrast on surface resistance

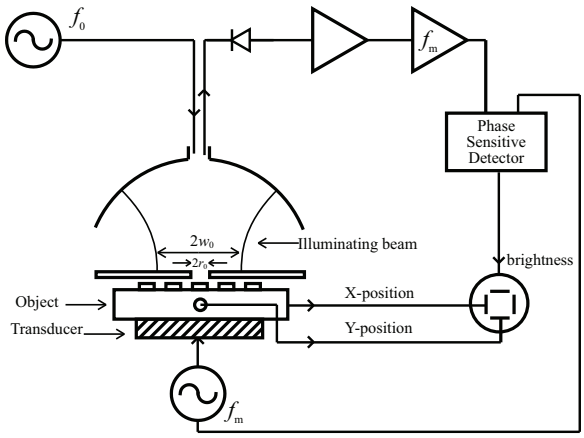


Figure 1.1: A 10 GHz oscillator excites a hemispherical resonator which is loaded by a sub-wavelength aperture of diameter 1.5 mm. [Taken from [6]].

of films in the range 0.5 Ω−cm and 100 Ω−cm with 1 mm spatial resolution.

Finally, in 1972 NFMM was introduced to a wider audience with an article in the journal Nature, by Ash and Nicholls [6] in which they demonstrated 0.5 mm spatial resolution with a sub-wavelength aperture in a 10 GHz resonator, a true decoupling of resolution and wavelength in a true non-contact scan. Their apparatus is reproduced in Fig. 1.1 for reference. They were able to differentiate dielectric permittivities between $\epsilon_r = 2$ and $\epsilon_r = 3$. This work is often cited as the first NFMM experiment, but as we have shown it was a more incremental development. Regardless, it is clear the field grew rapidly after this demonstration. Because Ash and Nicholls did introduce several critical elements in the system architecture we examine it in more detail. First, they included a high Q resonator, and they scanned a sub-wavelength aperture in the resonator over the sample without contact. Second, they employed phase sensitive detection (discussed in detail in Chapter 3) and modulation of a piezoelectric motion stage. And third, they monitored shifts in the resonant frequency and Q of the resonator. This architecture has been nearly universally copied as the best method for NFMM. Finally, Ash and Nicholls introduced the concept of perturbation theory as a means for quantifying measurements. We will demonstrate this later in chapter 2.

After the publication by Ash and Nicholls in Nature we no longer follow a linear timeline as many groups began work in parallel. Here we will aim to highlight the major developments

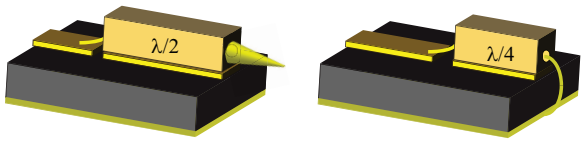


Figure 1.2: The electric-field probe is a sharp tip which looks like a small capacitor coupled to the DUT. It is close to an open circuit and thus peaks the voltage at the end. The magnetic field probe is a small wire connecting the resonator to the ground plane and looks like an inductor coupled to the DUT. It is close to a short circuit and thus peaks the current at the end.

and trends. In 1984 Massey formulated a theoretical basis for high spatial resolution with sub-wavelength apertures based upon an angular spectrum expansion of the near-field [37]. Then, by calculating the energy in a resonator-based NFMM probe, and the energy stored in the sample, Matey and Blanc showed that sensitivity was directly proportional to Q [38]. Today high Q resonators are almost exclusively used for NFMM.

Aside from the traditional aperture-based probe, Epstein et. al. applied an open-ended coaxial probe [39] similar to Bryant and Gunn's guided wave probe in 1965. In the same year, Gutmann et. al. demonstrated for the first time the difference between an electric field probe and a magnetic field probe [40]. This was later examined more rigorously and popularized by Tabib-Azar et. al. [41]. The electric field probe is a sharp metallic tip, while the magnetic field probe is a small wire loop, as shown in Fig. 1.2.

One further contribution has been that of topography-sample differentiation. It is unavoidable that the response of an NFMM is not only a function of the device under test (DUT), but also of the separation of the probe and the DUT [41]. In the simplest case, a strong $1/r$ response is expected from a parallel plate capacitor. This signal, if not controlled or anticipated can easily overwhelm the signal of interest. In 1981, at IBM laboratories, Binning and Rohrer demonstrated the STM, and by 1986 had perfected the AFM. Both methods of SPM proved to be extremely effective means of nanometer height control. Nearly two decades later, a hybrid AFM-NFMM was developed [8] such that microwave measurements could be decoupled from topography. The advantage of AFM-NFMM is two-fold: (1) the tip-DUT signal is decoupled from the topography signal with nanometer accuracy, and (2) because of the extremely close scan heights the evanescent NFMM signal is strongly coupled to the DUT. The problem with AFM-NFMM is also two-fold:

(1) it requires extremely planar samples, and (2) scan areas are at most 100 μm [42]. There are two alternative methods for compensating for topography in a sample. One can measure the sample surface at various points along the extents of the scan area and then compensate for the approximate DUT topography as a sloped plane. This method is fast and capable of high topography DUTs, but its crude approximation is only effective for larger DUTs. Another method is to employ an optical height profilometer and measure the topography at each point along the scan, then compensate for any topographical change. In this work, because we aim for simplicity, speed, large vertical range, and large scan area, we employ both of the alternative methods.

1.5 The need for NFMM

With an ever increasing number of SPMs it is important to ask what unique capability the scanning NFMM offers. In brief, the historical overview in Section 1.4 indicates the following:

(1) Spatial resolution on the order of most other SPMs.

(2) Sensitivity to constitutive parameters (i.e., complex ϵ, μ).

(3) Sensitivity to surface topography and DUT volume.

(4) Characterization of DUTs parameters in an important frequency band (microwaves).

It is for these unique capabilities that NFMM has been added to the host of SPM diagnostic tools.

Table 1.1 compares the various SPM methods for spatial resolution, any requirements on the sample, and general considerations. It is shown that while AFM and STM are superb for atomic-scale spatial resolution they require flat, and in the case of STM conductive samples. In addition, the field of view (FOV) is limited to at most 100 μm [42], and like optical microscopy, they are only sensitive to parts of the sample very near the surface. We also note that x-ray microscopy can be destructive to modern electronics and biological samples. NFMM provides

Table 1.1: Scanning Probe Microscopy methodologies

Method	Parameters		Comment
	Resolution	Sample Requirement	
STM	atomic	conductive	small FOV
AFM	hard: atomic bio: 2 nm	flat	small FOV (<100 μm)
Optical	100's nm	reflective	surface only
SEM	100 nm	conductive	expensive
X-ray	5 μm, sub-micron (with synchrotron)	-	active (i.e., possibly destructive), not sensitive to surface, large, expensive
Near-field microwaves	sub-micron	-	large FOV, surface & volume (must de-tangle), electrically interactive

unique capabilities in the vast range of SPMs including: sub-micron spatial resolution, minimal sample preparation, large FOV, and sensitivity to the surface topography and sub-surface volume of a DUT.

Chapter 2

NFMM Concept and Probe Circuit

Alchemy is the art that separates what is useful from what is not by transforming it into its ultimate matter and essence.

—Philippus Aureolus Paracelsus

Contents

2.1 Overview

In this chapter we discuss sub-wavelength resolution, the measurement concept and definitions germane to this body of work. Next we discuss the design of a microwave probe circuit which is sensitive to small changes in a DUT, followed by modeling of the probe tip and probe circuit. Finally we apply analytical models and computer simulation to the electromagnetic coupling of the probe-tip and DUT to ensure our modeling is valid.

2.2 Sub-wavelength resolution in the Near-field

We begin with a theoretical basis for high spatial frequency electromagnetic waves based on the angular spectrum expansion [37, 43]. The angular spectrum expansion is a decomposition of an arbitrary field distribution into propagating and non-propagating (or evanescent) plane waves at various angles. As discussed in chapter 1, a near-field probe can take the form of a sub-wavelength aperture or a sub-wavelength tip. Here we consider an infinite conducting plane with a square aperture of dimension $l \times l$, where $l << \lambda$, as shown in Fig. 2.1. We would like to solve for the field at an arbitrary value of z, given the specified field in the aperture at $z = 0$.

Assume a plane wave normally incident from the left ($z < 0$) illuminates the aperture creating a tangential electric field $\mathbf{E}_{\tan}|_{z=0}$. The electric field for $z > 0$ obeys the wave equation and by nature of the sub-wavelength extent of the aperture we approximate the aperture field as uniform:

$$\left(\nabla^2 + k^2\right)\mathbf{E} = 0 \ (z > 0) \tag{2.1}$$

$$\mathbf{E}_{\tan}|_{z=0} = \mathbf{E_0}\text{rect}(x/l)\text{rect}(y/l). \tag{2.2}$$

The second order partial differential equation (PDE) (2.1) can be simplified to a second order

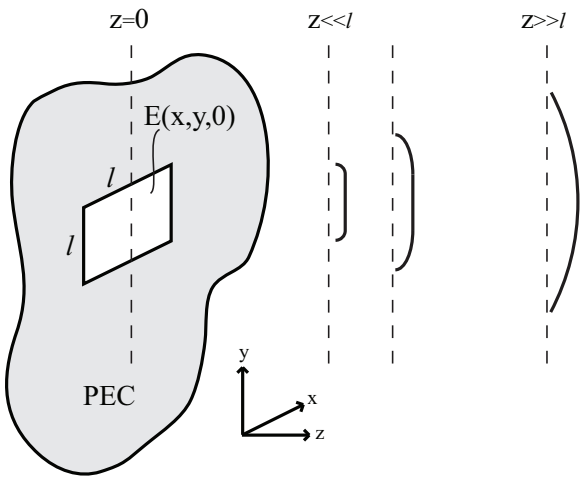


Figure 2.1: The near-field of an aperture caries with it spatial information on the order of the aperture size. This creates a sampling function of small spatial extent and allows the NFMM to measure a DUT in a small area.

ordinary differential equation (ODE) by application of the Fourier transform relation:

$$\widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};z\right) = \int\int \mathbf{E}(x,y,z)e^{-j2\frac{\pi}{\lambda}(\alpha x+\beta y)}dxdy \tag{2.3}$$

$$\mathbf{E}(x,y,z) = \int\int \widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};z\right)e^{j2\frac{\pi}{\lambda}(\alpha x+\beta y)}d\frac{\alpha}{\lambda}d\frac{\beta}{\lambda}, \tag{2.4}$$

where $\widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};z\right)$ is referred to as the angular spectrum of $\mathbf{E}(x,y,z)$, and α, β, and γ are direction cosines proportional to k_x, k_y, and k_z respectively. Equation (2.1) becomes

$$\left[\frac{d^2}{dz^2}+\left(\frac{2\pi}{\lambda}\right)^2\left(1-\alpha^2-\beta^2\right)\right]\widetilde{\widetilde{\mathbf{E}}}=0 \qquad (z>0), \tag{2.5}$$

with solution

$$\widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};z\right)=\widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};0\right)e^{-j\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z} \qquad (z>0), \tag{2.6}$$

where $\gamma=\sqrt{1-\alpha^2-\beta^2}$.

The $\widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};0\right)$ coefficient in (2.6) is the Fourier transform of the uniform aperture distribution in (2.2), thus

$$\widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};0\right)=l^2\mathbf{E_0}\text{sinc}\left(\frac{\alpha}{\lambda}l\right)\text{sinc}\left(\frac{\beta}{\lambda}l\right), \tag{2.7}$$

and from (2.6):

$$\widetilde{\widetilde{\mathbf{E}}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};z\right)=l^2\mathbf{E_0}\text{sinc}\left(\frac{\alpha}{\lambda}l\right)\text{sinc}\left(\frac{\beta}{\lambda}l\right)e^{-j\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z}. \tag{2.8}$$

Then by (2.4)

$$\mathbf{E}(x,y,z) = \int\int l^2 \mathbf{E_0} \text{sinc}\left(\frac{\alpha}{\lambda}l\right) \text{sinc}\left(\frac{\beta}{\lambda}l\right) e^{j\frac{2\pi}{\lambda}\left(\alpha x+\beta y-\sqrt{1-\alpha^2-\beta^2}z\right)} d\frac{\alpha}{\lambda} d\frac{\beta}{\lambda}. \tag{2.9}$$

The electric field in (2.9) is a superposition of plane waves with sinc weighting functions, some that propagate and others that decay exponentially. To ensure decaying exponentials along the z-axis

$$\gamma = \begin{cases} \sqrt{1-\alpha^2-\beta^2} & \text{if } \left(\alpha^2+\beta^2\right) < 1 \\ -j\sqrt{\alpha^2+\beta^2-1} & \text{if } \left(\alpha^2+\beta^2\right) > 1. \end{cases} \tag{2.10}$$

Therefore, waves in (2.9) with $\left(\alpha^2+\beta^2\right) < 1$ are propagating, while those with $\left(\alpha^2+\beta^2\right) > 1$ are exponentially decaying. The decaying waves are referred to as evanescent waves and are significant contributors to the total field only in the near-field of the probe-tip. When α and β are large, the wave vector in the transverse plane becomes larger than the propagating $z-$axis wave vector and it is this property which yields sub-wavelength lateral spatial resolution. It is important to note that the sinc weighting factor in (2.9), which corresponds to an $l \times l$ aperture, has significant content out to $1/l$ verifying that the amplitude of the evanescent waves corresponding to aperture-scale spatial frequencies are still significant. We note here that the low frequency plane waves with high spatial content decay along the z-axis but propagate in the $x-y$ plane at an angle from the positive x-axis of $\tan^{-1}\left(\frac{\beta}{\alpha}\right)$.

The result of this discussion is that it is possible to achieve sub-wavelength spatial resolution in the transverse plane, so long as the DUT is placed close to the sub-wavelength aperture ($z \leq l$). Figure 2.1 shows a conceptual diagram of a propagating field with small spatial extent for $z \ll l$. As z increases the field spreads to become a spherical wave in the far-field. The spatially confined field distribution near the aperture is effectively a sampling function of small spatial extent. In such a way spatial resolution on the order of the aperture dimensions is possible regardless of the operating wavelength.

2.3 Measurement Concept

The sub-wavelength aperture discussed in the previous section is the most obvious example of achieving sub-wavelength resolution through the near-field. However, an inverse structure results in much the same field localization. In contrast to the uniform distribution over the aperture extent, a sub-wavelength tip is modeled as a concentrated potential (e.g., a rectangular tip would have a rectangle-function boundary condition) and a similar resolution is achieved. In this work we have used a tip instead of an aperture because it is readily available from manufacturers and allows for high vertical aspect ratio measurements, one of the goals of this work.

Figure 2.2 shows the concept of an NFMM measurement. The probe circuit is generally described as a critically coupled resonator (f_0, Q), loaded by a near-field probe tip of characteristic diameter d_{tip}, maintained at a height h close to the surface of the DUT. The DUT is scanned beneath the probe tip by electromechanical and piezoelectric motion stages. Variations in the composition and structure of the DUT over the scan area cause changes in the near-field coupling of the tip and DUT which result in a change in the resonant response of the high-Q probe circuit. By monitoring the changing response of the probe circuit (Δf_0 and ΔQ, or $\Delta|S_{11}|$ at a fixed frequency f_{op}), variation in the DUT is mapped to contrast in the measurement.

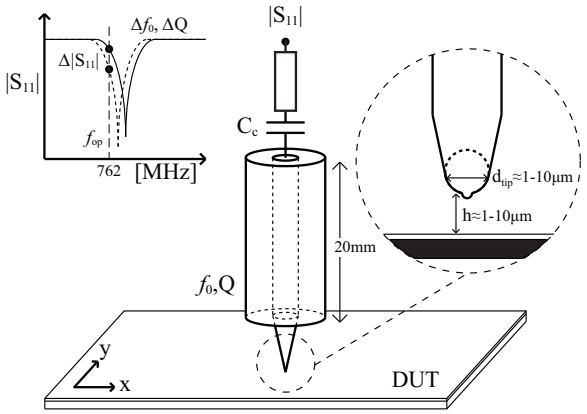


Figure 2.2: The NFMM probe circuit is a critically coupled resonator (f_0, Q), loaded by a near-field probe tip (inset–right) of characteristic diameter d_{tip}, at a height h above the surface of the sample. The probe tip is modeled as a shaft region, a tapered region, a spherical tip region, and a small imperfection at the apex of the tip. The reflection coefficient of the probe circuit (inset–left) is measured as the sample is scanned beneath the probe tip. Changes in the DUT cause changes in the probe response and result in contrast in the measurement.

The near-field probe tip is described by a straight cylindrical shaft which tapers to a spherical tip of diameter $d_{\text{tip}} \approx 1 - 10\,\mu\text{m}$. Imperfections on the tip are modeled as hemispherical bumps ($r \ll d_{\text{tip}}$) at the apex of the spherical tip. The tip is usually tungsten for its durability and soldered to the high-Q resonator of typical length 10–20 mm.

In a typical NFMM measurement the DUT is scanned to an (x, y) location while maintaining $h \approx d_{\text{tip}}$. The DUT and tip form a coupled electromagnetic system in which the capacitance is dictated by (2.34), and resistance is contributed by losses in the tungsten tip and/or the DUT. This complex impedance loads the resonator which causes its resonant frequency f_0 and Q to shift [44]. An RF excitation reflects off the resonator and the response is recorded for that location. The DUT is scanned to a new location and the process is repeated. The result is a multi-dimensional record (2D image, 1D line, etc.) proportional to the reflection coefficient of the probe circuit as a function of spatial coordinates. With a proper probe circuit model (discussed in section 2.5.2) and tip model (discussed in section 2.5.3) DUT parameters of interest can then be extracted from the dataset of reflection coefficients.

2.4 Definitions

For the purposes of this work we define certain terms that are referred to repeatedly throughout the text and describe various elements of the NFMM system. Details of each element are discussed in more detail in following chapters but for now we provide an overview. The numbers in Figures 2.3a, 2.3b, and 2.3c correspond to:

(1) **DUT**: the device under test, also referred to as the sample.

(2) **Probe Tip**: a sufficiently small structure with characteristic dimension much less than an operating wavelength which is used to couple to a localized region of the DUT. The electromagnetic coupling of the probe tip and DUT form a complex impedance which loads the probe circuit. In this work, the probe tip is a sharp tungsten tip with a cylindrical shaft diameter of approximately $0.5 - 1$ mm and tapering to a tip of diameter $d_{\text{tip}} = 1 - 20\,\mu\text{m}$.

See section 2.6.

(3) **Tip-DUT System**: the probe tip (2) and DUT (1) form a coupled electromagnetic system, shown in a dashed square in Figures 2.3b and 2.3c. In the quasi-static limit, it can be viewed as a complex impedance. See section 2.5.3.

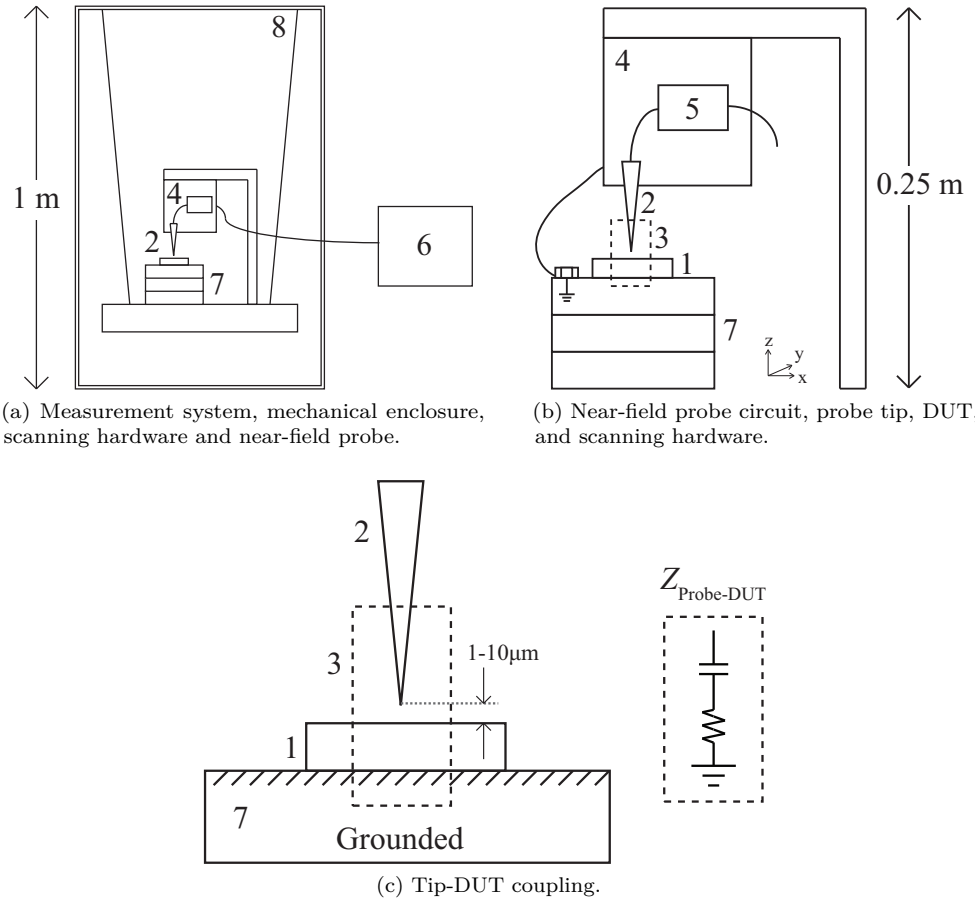
(4) **Probe Circuit**: a microwave circuit which serves as an interface between the probe tip and the readout instrument. It should enhance the small changes in the tip-DUT coupling impedance. In this work the probe circuit is a surface-mount microwave resonator of resonant frequency around 1 GHz. It is hybridized with microstrip transmission lines on a Rogers 4003c low-loss substrate. Together, the probe tip and the probe circuit are referred to as the probe. The probe circuit is held fixed above the DUT by an inverted L-shaped metallic fixture. See section 2.5.

(5) **Resonator**: we often refer to just the resonator of the probe circuit as it is the main functional element in the probe circuit. The resonator is critically coupled to the characteristic impedance of the instrument (6). It is a half-wavelength open-circuit ceramic filled rectangular coaxial transmission line with dielectric constant $\epsilon_r = 90$. See section 2.5.

(6) **Measurement System**: also referred to as the instrument, operates at microwave frequencies and is designed to measure the complex impedance of the probe circuit, similar to one-port of a traditional vector network analyzer (VNA), or reflectometer. In this work we refer to the instrument as the Lock-in Vector Near-field Probe (LVNP). See section 3.2.1.

(7) **Scan Assembly**: the DUT is scanned beneath the fixed probe with micromechanical and piezoelectric motion stages which comprise the scan assembly. The full assembly includes a stack of three mechanical stages with 1 μm step sizes for motion along each of the x, y, and z axes. At the top of the stack is a piezoelectric motion stage that acts only in the vertical or z-axis with sub-nm step sizes. The DUT platform on the piezo is connected to the RF ground of the microwave probe circuit (4) through a short length of wire. See section 3.3.3.

(8) **Mechanical System**: an acoustic and vibration isolation chamber in which the scan assembly, DUT, and probe sit. It is comprised of an acoustically insulated chamber and a 30 kg slab of granite suspended by a bungee-cord on each of four corners which forms a mechanical low pass filter. Cables pass through openings in the mechanical system enclosure to mate the instrument to the probe. See section 3.3.3.



(a) Measurement system, mechanical enclosure, scanning hardware and near-field probe.

(b) Near-field probe circuit, probe tip, DUT, and scanning hardware.

(c) Tip-DUT coupling.

Figure 2.3: The numbers in these figures correspond to those listed in section 2.4. (a) depicts the entire probe tip, probe circuit, DUT, and scanning hardware in the acoustic and vibration isolation chamber. (b) shows just the probe circuit and probe tip suspended above the DUT on the scanning hardware. A ground connect wire is shown as well. (c) shows detail of the tip-DUT coupling.

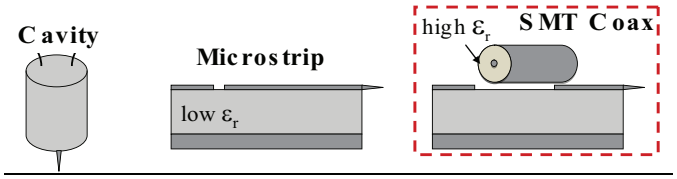
2.5 Probe Circuit

As described in Chapter 1, aperture-based probes and air-filled cavity resonators were used almost exclusively at the beginning of NFMM development. The advantages of a sub-wavelength aperture in a cavity sidewall are good isolation of the fields, and high Q. However, the air-filled cavities are large and the aperture-based probes prevent scanning samples with any significant topography.

The large air-filled cavity resonator can be replaced by a microstrip resonator which has the advantage of simplicity and flexibility but at the cost of greatly diminished Q. A microstrip probe circuit requires the aperture be replaced by a probe tip which allows for more flexible scanning.

A third option strikes a balance between the two. The shielded air-filled cavity resonator is replaced by a high-dielectric ($\epsilon_r = 90$), low-loss ceramic coaxial transmission line resonator. The high dielectric keeps the fields confined to the inside of the resonant structure allowing for more controlled scanning (i.e., only fields from the tip interact with the sample) and a high quality factor (over 1000) despite the open-ended structure. Finally, the transmission line resonator is a surface-mount component so it can be easily integrated with a hybrid circuit, therefore providing the simplicity and flexibility of the microstrip probe circuit with the high Q of the air-filled cavity. The probe is a tip instead of an aperture. These three options are compared in Fig. 2.4, with the result that in this work we use the surface mount coaxial transmission line resonator for its high performance, flexibility, and ease of integration.

Figure 2.5 (left) shows a photograph of the microstrip implementation of the probe circuit built on Rogers 4003c substrate. A series capacitor C_c couples the incident RF power into a surface-mount rectangular coaxial transmission line resonator (SkyWorks SR9000EPHY790) of resonant frequency $f_0 = 790\,\text{MHz}$, with a quality factor $Q \approx 1400$. The $\lambda/2$ resonator is loaded by a sharp metallic near-field probe tip with a characteristic tip diameter d_{tip} on the order of 1–10 μm, to be scanned at a height $h \approx d_{\text{tip}}$ above the surface of the DUT. The probe tip has four distinct regions (Fig. 2.2, inset–right): a long straight shaft which is electrically connected to the center conductor of the probe circuit resonator through a length of transmission line, a



Circuit Topology	Q	Size	Tip Integration	Topology
Cavity	~1700	Large	Difficult; Isolated	Coax
Microstrip	~100	Medium	Simple; Un-shielded	Planar
SMT Coax	~1600	Small	Simple; Partially Shielded	Planar

Figure 2.4: A large air-filled cavity provides high Q-factor but is large, difficult to work with, and hard to integrate with various probe tips. A microstrip resonator is of medium size, is flexible, and easy to work with however suffers from low Q-factor. A ceramic dielectric ($\epsilon_r = 90$) coaxial resonator is surface mount so it is easy to work with, but also provides high Q-factor. The probe circuit in this work uses the high permittivity coaxial resonators.

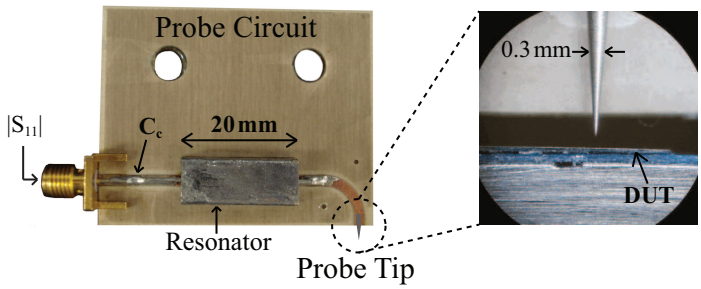


Figure 2.5: The near-field probe circuit (left) is comprised of a capacitively coupled 790 MHz resonator, loaded by the near-field probe tip (right).

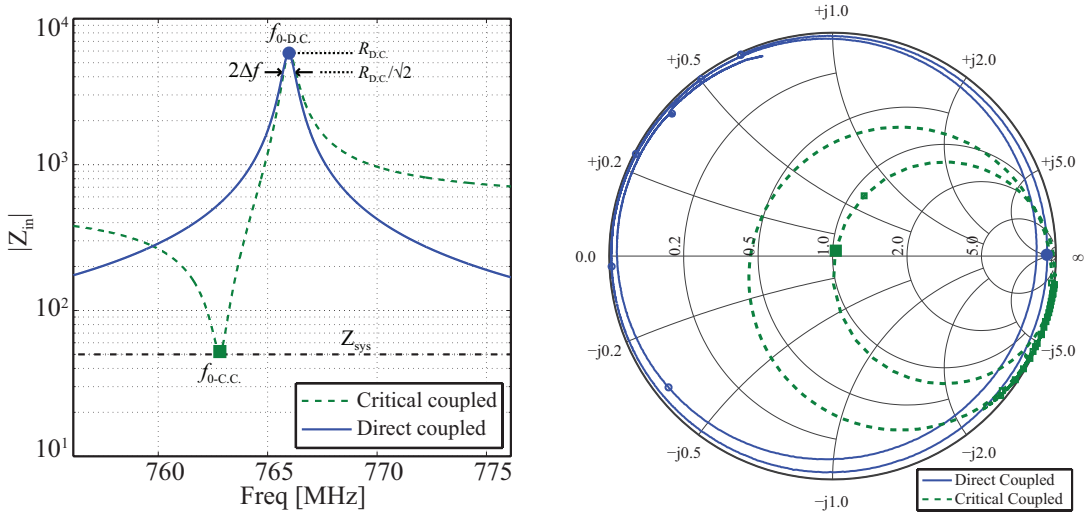
tapered transition region to the spherical tip region, with a small imperfection at the apex of the tip modeled as a spherical bump [45].

2.5.1 Resonator Coupling & Q-factor

It is simpler to measure a small perturbation to a small signal than to a large signal as it requires less instantaneous dynamic range. Since we are measuring small perturbations to the reflection coefficient we desire the resonator be impedance matched [46] to reduce the background signal. This condition is referred to as critically coupled. In effect, an impedance transformation converts the impedance seen at resonance to the system characteristic impedance, in our case 50 Ω. The input impedance of an open-circuited half-wavelength transmission line in the critically-coupled and the direct-coupled case is compared in Fig. 2.6, where a series coupling capacitor C_c, converts the high input impedance of the parallel resonance to a matched series resonance. The solid blue circle indicates the impedance at resonant frequency $f_{0\text{-D.C.}}$, for a direct coupled probe circuit while the solid green square indicates the impedance at resonant frequency $f_{0\text{-C.C.}}$, for a critically coupled probe circuit. Notice that the green square is almost exactly at the center of the smith chart indicating a near-perfect match at resonance.

The quality factor (Q) of a resonator is defined as the ratio of average energy stored in the resonator to the energy lost in the resonator:

$$Q \equiv 2\pi \frac{W_m + W_e}{W_{\text{loss}}} = 2\pi \frac{W_m + W_e}{\frac{P_{\text{loss}}}{[\text{sec}]}} = 2\pi \frac{W_m + W_e}{P_{\text{loss}} \frac{1}{f}}, \tag{2.11}$$



(a) $|Z_{\text{in}}|$ (b) Z_{in}

Figure 2.6: The $\lambda/2$ open-circuit resonator is a parallel resonance and thus has a high input impedance at resonance (almost 6 kΩ). Therefore, it must be critically coupled to reduce the required dynamic range in the reflection measurement. A series coupling capacitor C_c serves as an impedance inverter changing the parallel resonance to a series resonance and matching the impedance at resonance to the system impedance of 50 Ω.

and at the resonant frequency f_0, $W_m = W_e$ such that

$$Q_0 = \omega_0 \frac{2W_m}{P_l}. \tag{2.12}$$

We will refer to Q_0 as Q where it is understood that it is defined around the resonant frequency and refers to the intrinsic Q of the resonator, or the unloaded quality factor (Q_U). It can be shown that [47]

$$Q = \frac{1}{\delta(p)} \sqrt{\frac{1-p}{p}}, \tag{2.13}$$

where p is the power fraction and $\delta(p) = \frac{2\Delta\omega}{\omega_0}$ is the fractional bandwidth at the power fraction p. While (2.13) defines the Q_U at any power level it is traditional to set $p = 0.5$:

$$Q = \frac{1}{\frac{2\Delta\omega_{0.5}}{\omega_0}} \sqrt{\frac{1-0.5}{0.5}} = \frac{\omega_0}{2\Delta\omega_{0.5}} = \frac{1}{BW}, \tag{2.14}$$

where $2\Delta\omega_{0.5}$ is the full-width bandwidth between half-power ($p = 0.5$) points around the resonant frequency and BW is the usual fractional bandwidth. In other words, a high-Q resonator will have a very narrow resonance while a low-Q resonator will have a broad resonance. From the perspective of NFMM, as will be discussed in more detail in section 2.5.2, it is advantageous to

have a narrow resonance in which impedance changes rapidly around the resonant frequency. The input impedance at resonance is purely real and the half-power point corresponds to

$$Z_{\text{in}} = \begin{cases} \frac{R}{\sqrt{2}} & \text{parallel resonance} \\ \sqrt{2}R & \text{series resonance,} \end{cases} \tag{2.15}$$

labeled in Fig. 2.6 for the direct coupled parallel resonance. In this work we always describe a resonator by Q_U and extract it by converting $|S_{11}|$ into $|Z_{\text{in}}|$ and applying (2.14) and (2.15).

Though a resonator is described by its intrinsic or unloaded quality factor, the necessary act of coupling energy into and out of the resonator introduces external loading which always broadens the resonance. The external circuit can be described by its own quality factor or external quality factor (Q_E) and then the loaded resonant circuit exhibits a loaded quality factor (Q_L) where,

$$Q_L = \frac{Q_U}{1+\beta} \tag{2.16}$$

$$Q_E = \frac{Q_U}{\beta}, \tag{2.17}$$

and $\beta = \frac{Q_U}{Q_E}$ is the coupling coefficient:

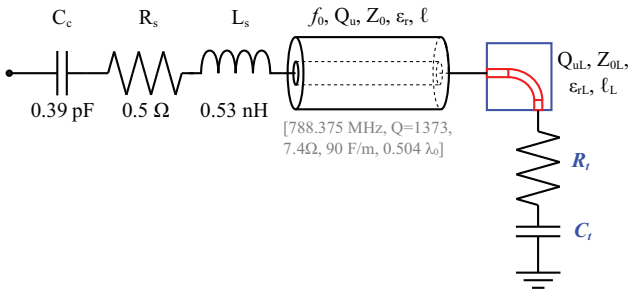
$$\begin{array}{ll} \beta < 1 & \text{under-coupled} \\ \beta = 1 & \text{critically-coupled} \\ \beta > 1 & \text{over-coupled.} \end{array} \tag{2.18}$$

Since we have critically coupled the resonator in the probe circuit, $\beta = 1$ and thus $Q_E = Q_U$ and $Q_L = \frac{Q_U}{2}$. We note that the critically coupled condition is tantamount to the maximum power transfer impedance match.

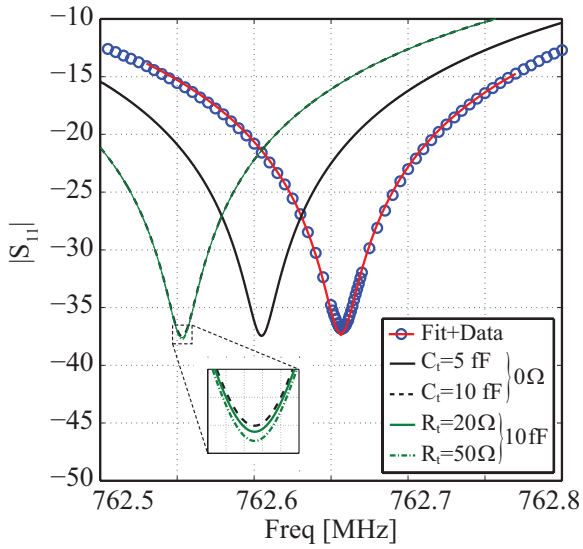
2.5.2 Equivalent Circuit Modeling

In order to work efficiently with the NFMM system and be able to predict trends we require an equivalent circuit for both the probe circuit and the probe tip-DUT system. Because the microwave probe circuit resonator is on the order of an operating wavelength we expect to include a distributed transmission line in the equivalent probe ciruit. However, the probe tip-DUT system is sub-wavelength and thus we would like to model its impedance as a lumped element.

Figure 2.7a shows an equivalent model of the probe circuit. The coaxial transmission line resonator is a half-wavelength open-circuit resonator, which has a high impedance at resonance. As discussed in section 2.5.1 it is critically coupled with a series capacitor C_c in order to map this resonance into the most sensitive region of the smith chart for reflection-type measurements [46]. A series resistor and inductor model losses and tab inductance respectively while a small microstrip transmission line bend models the microstrip circuit where the tip is attached to the coaxial resonator. The tip-DUT coupling is modeled as a complex impedance. The capacitive coupling of the tip to the DUT is represented by the lumped element capacitor C_t, while losses in the tip metal and the DUT are modeled by the lumped element R_t.



(a)



(b)

Figure 2.7: (a) An equivalent lumped-element and transmission line circuit model is used for fitting the response of the probe circuit to determine the tip-DUT coupling impedance. (b) Typical capacitive loading shifts the resonance down dramatically while typical resistive loading slightly decreases the minimum $|S_{11}|$.

Table 2.1: Probe circuit model parameters

Parameter	Value	Sweep Min	Sweep Max
C_c	0.38 pF	0.3 pF	0.7 pF
R_s	0.2 Ω	0 Ω	0.5 Ω
L_s	~ 0	0 nH	2 nH
Length	0.5 λ_0		
Q_U	1286	250	2000
Z_0	7.04 Ω		
f_0	765.9597 MHz		

As mentioned in section 2.5 the high dielectric constant of the probe circuit resonator confines the fields inside the resonator such that the probe circuit response is not a function of its surroundings, except for the tip-DUT loading. Therefore, to find appropriate values for the probe circuit model we fit the model to a measurement of the probe circuit in free space. This ensures the tip-DUT coupling does not effect the fitting process. Figure 2.7b shows the measured and fitted response of the probe circuit in free-space (i.e., tip-DUT coupling is negligible) with fit parameters shown in Tab. 2.1.

As the tip approaches the DUT, the loading increases and the resonance shifts. Referring again to Fig. 2.7b, as C_t loading increases to 5 and 10 fF the resonant frequency shifts down dramatically while increased R_t loading causes a small change in the minimum of $|S_{11}|$ (inset). It is important to note that the small change due to resistive coupling can only be measured with sufficient capacitive coupling (here $C_t = 10$ fF). This is equivalent to saying that the probe tip must be "close enough" to the DUT for a resistive measurement to be successful.

Now that we have an equivalent circuit we wish to study the parametrics of the probe circuit to see how sensitive it is to certain factors. For example, we would like to see how stable the coupling capacitor C_c must be in order to reliably attribute a changing signal to the DUT and not C_c. Some parameters, such as f_0 and length l predictably shift the frequency response of the probe model. However, C_c, R_s, L_s, and Q_U have more meaningful effects. Column 3 and 4 of Table 2.1 show a range of variation for these parameters, and Figures 2.8 through 2.10 show the response of the probe circuit model under these conditions.

Because the probe circuit model is a series circuit, the coupling capacitor and the series

parasitic resistance could be equivalently exchanged with the tip-DUT coupling impedance and thus any change in C_c or R_s are indistinguishable from a change in the tip-DUT coupling impedance. Figure 2.8 shows the extreme sensitivity of the resonant response of the probe circuit to the coupling capacitor and the parasitic resistance. Connecting and reconnecting the probe circuit must be done carefully and with a torque wrench in order to prevent changing values of R_s and C_c from corrupting meaningful comparison of independent measurements. Also, the critical coupling capacitor must be extremely stable. Initially we used a screw-tuned capacitor to achieve optimal critical coupling of the resonator. However, it proved to be too unstable and so we have opted for a fixed coupling capacitor to eliminate any issues with the coupling capacitance changing during a measurement. Similarly the series inductance will be susceptible to connection and disconnection, as well as flex in the cables as shown in Fig. 2.9. This value should be kept extremely stable as small changes in series inductance will significantly alter the resonance and prevent any repeatable extraction of tip-DUT impedance.

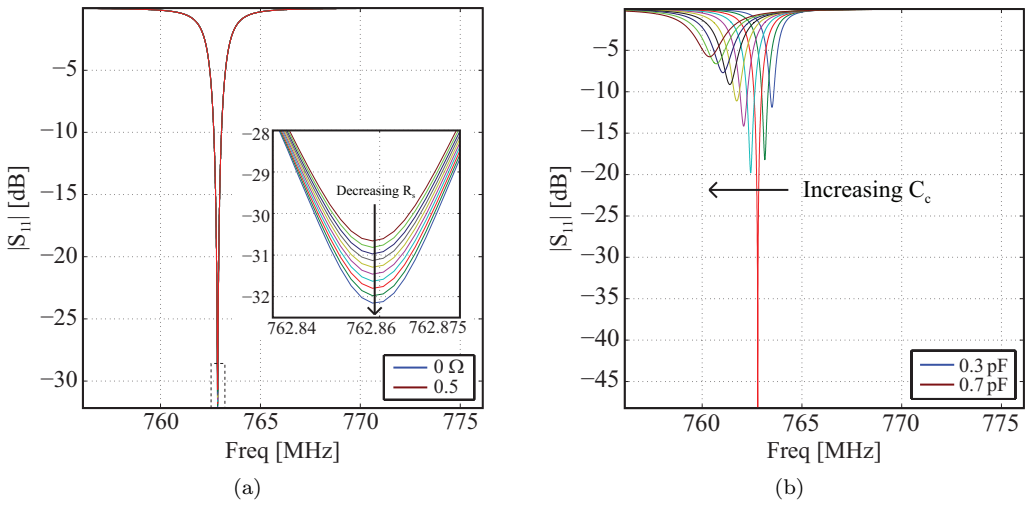


Figure 2.8: (a) Small changes in the parasitic series resistance are indistinguishable from tip-DUT losses. (b) Similarly, small changes in the series critical coupling capacitor are indistinguishable from tip-DUT capacitance coupling.

Figure 2.10 shows how variation in the Q of the probe circuit changes the resonant response. In practice we do not anticipate the resonator Q changing significantly during operation, however we include this plot to discuss the role of increasing Q. As we know from (2.14) Q is related to

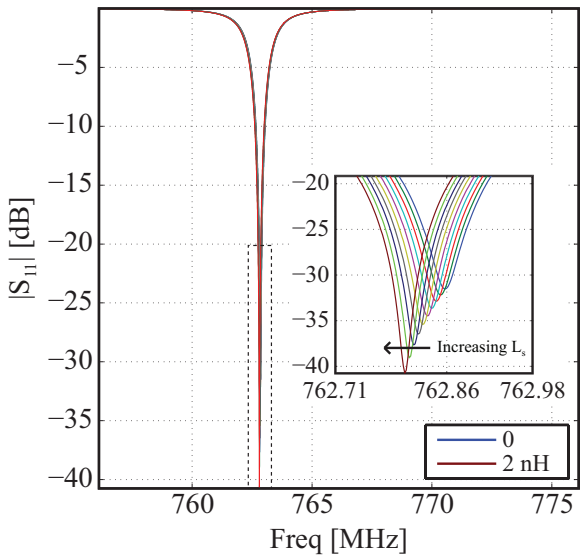


Figure 2.9: Small variations in series inductance shift the resonance and should be minimized. L_s can vary by connecting and disconnecting the probe circuit but otherwise will remain relatively stable.

the fractional bandwidth of a circuit around the resonant frequency. Increasing Q always narrows the resonance but it does not always deepen the resonance. The depth of the resonance is only a function of the real part of the input impedance at the resonant frequency (i.e., it is only a function of how well we have critically coupled the resonator). As Q diverges from its nominal value, critical coupling is no longer achieved. In this simulation the nominal Q is 1286. As Q increases from 200 to just below nominal ($Q = 1222$) the resonance narrows and approaches a better match (getting deeper) as indicated by solid traces. At $Q = 1286$ (where we have critically coupled the probe circuit) the resonance is narrower still and the match is best. However, as we continue to increase to $Q = 1417$, the response continues to narrow but the match begins to degrade from the critical coupled state as indicated by dashed traces. At $Q = 2000$ the response is the narrowest but the match is only -14 dB. Obviously if Q changes from its nominal value the resonator input impedance changes and we must recalculate the value of C_c to achieve critical coupling.

Referring again to the equivalent circuit in Fig. 2.7a and neglecting the parasitic elements and the explicit microstrip transmission line between the resonator and the probe tip, we can express

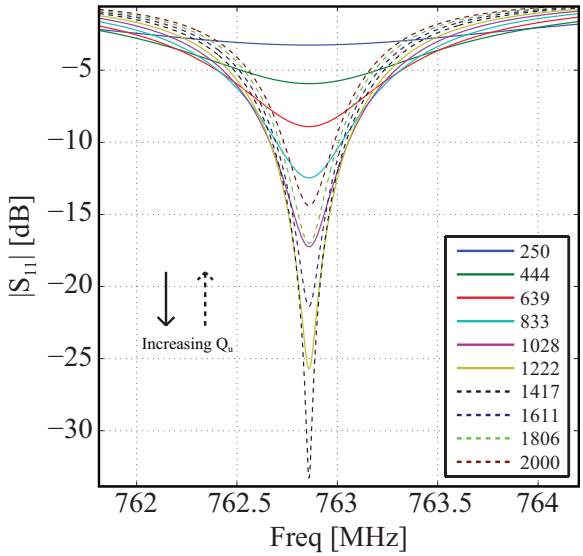


Figure 2.10: Q only fundamentally affects the width of the resonance. However, because the resonance is changing it also appears to have an effect on the input match. This is only because the probe circuit was critically coupled for the nominal values of the probe circuit and by changing Q the critical coupling is no longer tuned correctly, thus the depth of the resonance changes.

the input impedance in terms of the tip impedance Z_t and the quality factor of the resonator Q.

$$Z_t = R_t + \frac{1}{j\omega C_t}, \tag{2.19}$$

and

$$Z_{\text{probe}} = Z_0 \frac{Z_t + Z_0 \tanh\gamma l}{Z_0 + Z_t \tanh\gamma l}, \tag{2.20}$$

where $\gamma = \alpha + j\beta$.

With,

$$\beta = \frac{2\pi}{\lambda_g} \tag{2.21}$$

$$\alpha = \frac{\beta}{2Q}, \tag{2.22}$$

then

$$\gamma = \frac{2\pi}{\lambda_0}\left(j + \frac{1}{Q}\right), \tag{2.23}$$

and combining (2.19), (2.20), and (2.23) yields an explicit expression for Z_{probe} in terms of the tip impedance and the resonator Q. The limit of (2.23) as $Q \to \infty$ is $\gamma = j\beta$ and (2.20) becomes the familiar:

$$Z_{\text{probe}} = Z_0 \frac{Z_t + jZ_0 \tan\beta l}{Z_0 + jZ_t \tan\beta l}. \tag{2.24}$$

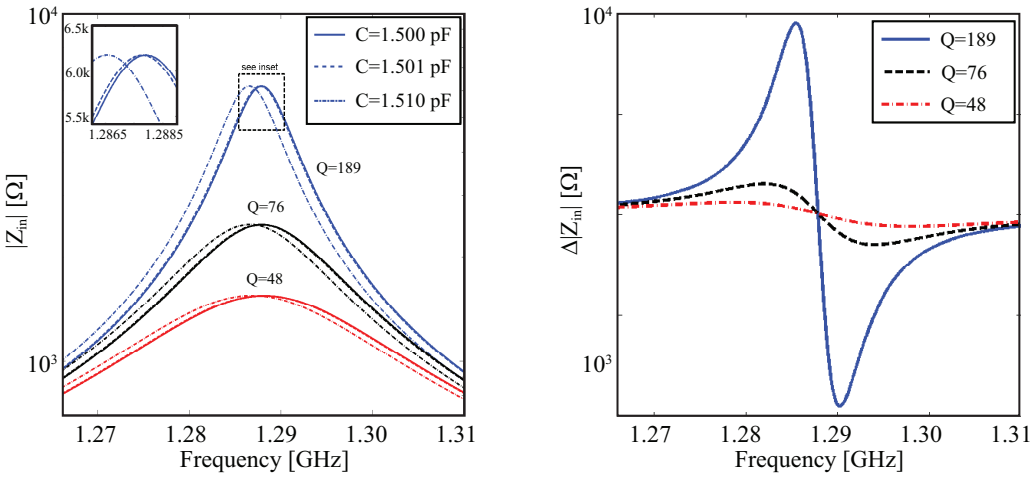
To demonstrate the dependence upon Q we fabricated three microstrip resonator probe circuits with measured Q_U of 48, 76, and 189 and compared their impedance response as load impedance (Z_t) was varied by a small amount. The measured frequency response of the probe circuits with a nominal 1.5 pF tip capacitance was imported into Applied Wave Research Microwave Office (MWO) where we then simulated the effect of small perturbations to the tip impedance of 1 and 10 fF. These small changes in capacitance cannot be realized in a packaged device but are seen in realistic NFMM measurements. The result is shown in Fig. 2.11 where the higher Q circuits yield a larger change in the input impedance of the resonator and thus a larger change in the measured signal. Figure 2.11a shows the input impedance for each resonator with nominal and perturbed loading, and Fig. 2.11b emphasizes the effect by plotting the difference between the nominal load of 1.5 pF and the perturbed load of 1.501 pF. The change in input impedance is enhanced by a high-Q resonator and therefore we conclude that a critically coupled, high Q probe circuit is the optimal NFMM circuit. Critical coupling impedance matches the probe circuit to the readout system (LVNP) such that the reflection coefficient is low (−30 to −40 dB), ensuring a small background signal which requires a lower dynamic range measurement. The high Q resonator means the input impedance of the probe circuit has a larger change per unit change of the probe-tip impedance.

2.5.3 Tip Modeling with Perturbation Theory

We would like to further justify our selection of modeling the tip-DUT system as a complex impedance as was suggested in Fig. 2.3c and 2.7a. We can motivate this by appealing to perturbation theory [44]. The entire resonant structure consists of the transmission line resonator, the probe tip, and the sample. The high dielectric constant of the resonator ensures that the energy inside the resonator is much greater than that around the tip and in the sample, therefore we can apply perturbation theory as in [19]. The shift in resonant frequency is

$$\frac{f - f_0}{f} = \frac{-\int\int\int_V \left(\Delta\epsilon \mathbf{E}\cdot\mathbf{E_0^*} + \Delta\mu \mathbf{H}\cdot\mathbf{H_0^*}\right) dV}{\int\int\int_V \left(\epsilon \mathbf{E}\cdot\mathbf{E_0^*} + \mu \mathbf{H}\cdot\mathbf{H_0^*}\right) dV}, \tag{2.25}$$

where subscript 0 represents the unperturbed case. The unperturbed energy is $W_0 = W_{\text{resonator}} + W_{\text{tip}} + W_{\text{surroundings}}$, but because the fields inside the cavity are approximately unaffected by



(a) $|Z_{\text{in}}|$ (b) $\Delta|Z_{\text{in}}|$

Figure 2.11: (a) Microstrip circuits with $Q_u = 48$, 76, and 189 were built and measured with a 1.5 pF capacitor at the end. In MWO a 1 and 10 fF variation was simulated. (b) The difference between the 1.5 pF and the 1.501 pF responses are plotted over frequency for each of the three resonators. A higher Q-factor results in a larger change in input impedance.

the tip and surroundings (by virtue of the high dielectric constant and the small volume of the tip-DUT system), when the sample is introduced only the changes in $W_{\text{tip}} + W_{\text{surroundings}}$ need be considered. For non-magnetic samples, where $\Delta\mu = 0$, the perturbed energy is:

$$\delta W = -\int\int\int_V \Delta\epsilon \mathbf{E} \cdot \mathbf{E}_\mathbf{0}^* dv. \tag{2.26}$$

Finally, because the denominator of (2.25) remains approximately constant for perturbation analysis, then it is only the numerator, which we have simplified to (2.26), that is directly proportional to the change in resonant frequency:

$$\frac{f - f_0}{f} = k\delta W, \tag{2.27}$$

where $k = \left[\int\int\int_V (\epsilon \mathbf{E} \cdot \mathbf{E}_\mathbf{0}^* + \mu \mathbf{H} \cdot \mathbf{H}_\mathbf{0}^*)\right]^{-1} = \text{const}$. In other words, the shift in resonant frequency is directly proportional to the shift in energy stored by the tip-sample interaction.

Since the quasi-static energy stored in a capacitor is

$$W_{qs} = \frac{1}{2} C V^2, \tag{2.28}$$

and C is a function of ϵ_{eff}, then by modeling the tip-sample interaction as a capacitor, we can say

that $\delta W = W_{qs}$, and the shift in resonant frequency is proportional to the change in tip-sample energy, or to the tip-sample capacitance. Since capacitance is a function of the effective dielectric constant of the tip-DUT system, this is the relationship we exploit in NFMM measurements to correlate measured reflections with variation in the composition of the DUT.

2.6 Tip Modeling

The probe tip is, in general, a complicated structure as depicted in the inset-right of Fig. 2.2. We model the tip as a straight cylindrical shaft with a tapered conical section which transitions to a spherical tip. At the apex of the tip is a possible imperfection modeled as a hemisphere. The full model of the tip must be simulated as it is not amenable to analytical solution. However various approximations on the tip geometry do yield analytical solutions.

2.6.1 Analytical

There are many analytical approaches one can take but all of them rely upon two simplifying assumptions: (1) the salient features of the tip can be approximated by a simple geometry such as a sphere or an axially rotated hyperbola, and (2) the problem is approximately electrostatic. The second assumption is easily justified by the fact that the entire tip structure is often 2-3 orders of magnitude smaller than a wavelength. Our aim is to find a simple analytical expression that approximately models the most important features of the tip-DUT coupling. We examine three analytical methods.

The first analytical method is based upon casting the problem in prolate spheroidal coordinates [48] such that the sample and tip lie on surfaces of constant η, as shown in Fig. 2.12. The method of images is employed to enforce equipotentiality at $\eta = 0, \eta_0$, and both point charges and segments of charge are used (Q_n and Q_m respectively). From [49], the potential is found as:

$$V(\rho, z) = \sum_{n=1}^{N} \left[V_{cn}(\rho, z; z_n) - V_{cn}(\rho, z; -z_n) \right] + \sum_{m=1}^{M} \left[V_{sm}(\rho, z; z_m) - V_{sm}(\rho, z; -z_m) \right], \quad (2.29)$$

where V_{cn}s correspond to discrete charges Q_n and V_{sm}s correspond to segments of charge density

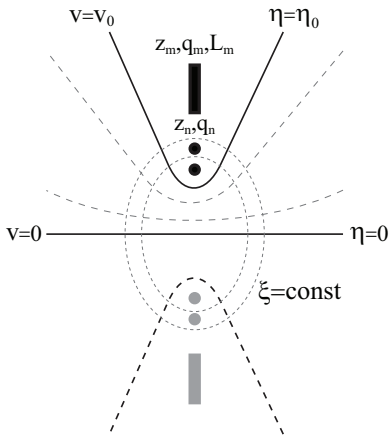


Figure 2.12: The probe tip over a ground plane can be modeled with image theory in the prolate spheroidal coordinate system in which the tip and ground conductors are on planes of constant ρ. The boundary conditions are satisfied by placing point charges and charge segments along the axis of rotation.

$Q_m/(2L_m)$. There are five unknowns in this solution: point charge strength Q_n, length of charge segments $2L_m$ and total strength of charge segments Q_m, as well as the position of each charge and segment (ρ, z). This problem can be solved for simplified geometries (e.g., spheroidal and hyperbolic tips) and has proven to be very accurate, however it still does not capture the effect of a cylindrical shaft and results in a less intuitive expression than we would like with too many unknowns.

A second possible analytical method is based upon a simplified multipole expansion technique. Figure 2.13 shows a conducting sphere of diameter d_{tip} at a height h over an infinite planar conducting ground plane. For simplicity we would like to truncate the multipole series to just two sources, the monopole (point charge) and the dipole. However, because of the strong fields when the sphere is close to the plane it is found that the two-term multipole expansion is only accurate for $h > d$. The near-field is strongest within this range and thus the two-term multipole expansion will not suffice.

The final analytical method, which we will work out in detail, is based upon repeated method of images [50] which can be applied in mixed cartesian/cylindrical coordinates as shown in Fig. 2.14. We replace the spherical conductor with a point charge Q_0 at the center of the sphere at $z = h$. In order to enforce equipotentiality at the $z = 0$ plane we require $-Q_0$ at $z' = -h$.

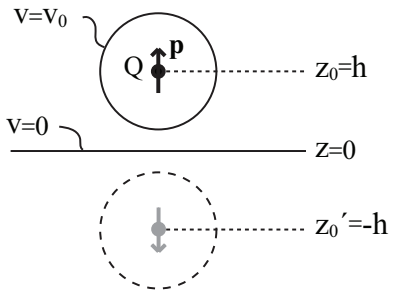


Figure 2.13: The probe tip over a ground plane can be modeled with image theory in the prolate spheroidal coordinate system in which the tip and ground conductors are on planes of constant ρ. The boundary conditions are satisfied by placing point charges and charge segments along the axis of rotation.

However, this upsets the potential on the spherical surface so we now require a second point charge Q_1 to bring the sphere back to equipotentiality. This too is imaged over the $z = 0$ plane and the process continues. The result is a series of point charges Q_i at $z = h - d_i$:

$$Q_i = -\frac{a}{d_i} Q_0 \tag{2.30}$$

$$d_i = \frac{a^2}{2d_{i-1}}, \tag{2.31}$$

where $d_0 = h$, and capacitance is defined as

$$C_{\text{tip}} = \frac{Q_{\text{tot}}}{V_0}. \tag{2.32}$$

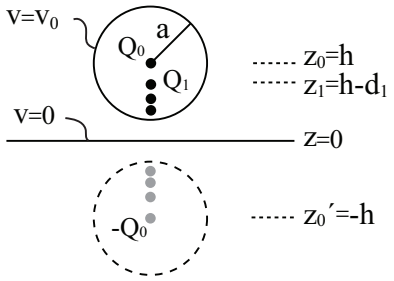


Figure 2.14: The probe tip over a ground plane can be modeled with image theory in a mixed spherical-cartesian coordinate system by enforcing equipotential surface at every point along a spherical tip and along a planar ground plane. The result is an infinite series that provides good accuracy with the first three terms.

The total charge Q_{tot} is the sum of all Q_is on just one electrode (i.e., only the positive or only the negative electrode, not both). V_0 is the electrostatic potential due to the entire ensemble of

charges but because of symmetry all charge pairs (Q_i and $-Q_i$) cancel leaving only Q_0. Therefore

$$C = \frac{Q_{\text{tot}}}{\frac{Q_0}{4\pi\epsilon_0 a}} = \frac{Q_0 + Q_1 + Q_2 + Q_3 + \cdots}{Q_0/4\pi\epsilon_0 a} \tag{2.33}$$

$$C = 4\pi\epsilon_0 a \left(1 + \alpha + \frac{\alpha^2}{1-\alpha^2} + \frac{\alpha^3}{(1-\alpha^2)(1-\frac{\alpha^3}{1-\alpha^3})} + \ldots\right), \tag{2.34}$$

where $\alpha = \frac{d_{\text{tip}}}{4(h+d_{\text{tip}}/2)}$. This expression is simple and the three term approximation results in little error over a broad range of heights.

The spherical capacitor over a ground plane is approximately equal to a circular parallel plate capacitor of diameter $2d_{\text{tip}}$ over a very small range around $h = d_{\text{tip}}$. Figure 2.15 shows the four term approximation to (2.34) for $a = [1, 5, 10]\,\mu$m, as a function of height with typical expected values in the range of 1 fF. An important difference between (2.34) and the parallel plate capacitor is the saturation at large (dashed traces) and small separation distances. That is,

$$\lim_{(h-a)\to\infty} C = 4\pi\epsilon_0 a = C_\infty \tag{2.35}$$

$$\lim_{(h-a)\to 0} C = 8\pi\epsilon_0 a = 2C_\infty. \tag{2.36}$$

In other words, as height increases the capacitance approaches that of a sphere in free space. As the tip approaches the conducting plane the capacitance approaches twice that of a sphere in free space.

2.6.2 Full-wave Simulation

While the analytical methods are good for insight into trends and functional dependencies (e.g., $1/r$, ln, power series, etc.) of tip capacitance, we require computer simulations in order to consider losses, to account for the full geometry of the tip, and to study depth penetration. In this section we perform full-wave simulations with Ansys HFSS$^{\text{TM}}$ (HFSS) and 2D quasi-static finite element method (FEM) simulations with an open-source software called Finite Elements Method for Magnetics (FEMM). HFSS will predict losses and depth penetration but involves a large computational domain. The 2D quasi-static simulation is sufficient for studying capacitive coupling between a complicated tip geometry and a sample and provides rapid

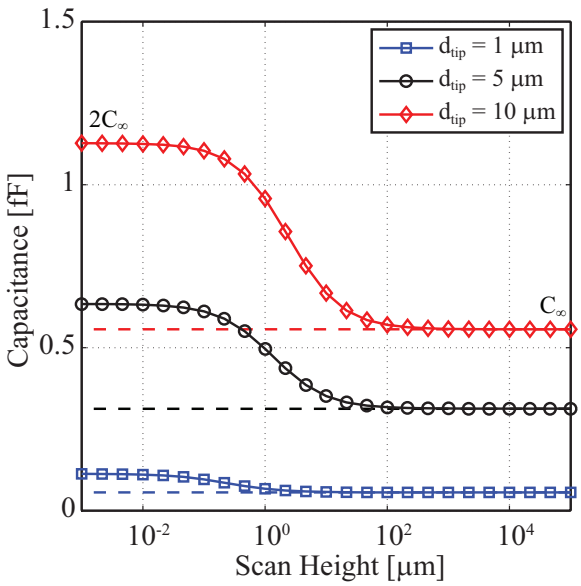


Figure 2.15: Tip capacitance calculated for a sphere over a ground plane varies from C_∞ far from the plane to $2C_\infty$ close to the plane. The region of rapid change is heights approximately equal to the tip radius.

convergence and a low memory footprint. In addition, we can automate FEMM from MathWorks MATLAB® (MATLAB) which is useful for fitting data, as will be shown in Chapter 4.

Figure 2.16 shows the concentration of the electromagnetic field in a 2.25 μm slab of copper beneath a 1 μm probe tip operating at 1, 3, and 4 GHz as simulated in HFSS. As expected the fields concentrate laterally around the diameter of the tip, but their depth penetration seems to be inversely related to frequency, similar to the concept of skin depth for TEM plane waves. To derive the skin depth for a good conductor we must introduce the concept of conductivity (σ) into Maxwell's equations:

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{2.37}$$

$$\nabla \times \mathbf{H} = \sigma\mathbf{E} + j\omega\epsilon\mathbf{E}. \tag{2.38}$$

From (2.37) and (2.38) the wave equation for a conductive medium is

$$\left[\nabla^2 - j\omega\mu\left(\sigma + j\omega\epsilon\right)\right]\mathbf{E} = 0 \tag{2.39}$$

$$\left[\nabla^2 - \gamma^2\right]\mathbf{E} = 0, \tag{2.40}$$

where $\gamma = \sqrt{j\omega\mu\left(\sigma + j\omega\epsilon\right)}$ is the complex propagation coefficient. γ can also be expressed by its

real and imaginary parts as $\gamma = \alpha + j\beta$. For a good conductor $\sigma \gg \omega\epsilon$ and

$$\gamma = \sqrt{j\omega\mu\left(\sigma + j\omega\epsilon\right)} \simeq \sqrt{j\omega\mu\sigma} \tag{2.41}$$

$$\gamma \simeq \sqrt{\frac{\omega\mu\sigma}{2}\left(1+j\right)^2} \tag{2.42}$$

$$\gamma \simeq \sqrt{\frac{\omega\mu\sigma}{2}}\left(1+j\right) = \alpha + j\beta \tag{2.43}$$

and therefore

$$\alpha = \beta = \sqrt{\frac{\omega\mu\sigma}{2}}. \tag{2.44}$$

The distance at which a plane wave is attenuated to its $1/e$ point is called the skin depth δ:

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu\sigma}} = \sqrt{\frac{\rho}{\pi f\mu}}, \tag{2.45}$$

where f is the frequency of the plane wave, σ is the bulk conductivity, $\rho = 1/\sigma$ is the bulk resistivity, and $\mu = \mu_r\mu_0$ is the material permeability. As conductivity increases skin depth decreases and approaches zero for a perfect electric conductor (PEC). Though it is hard to tell the precise functional dependence of the field penetration depths in Fig. 2.16 on frequency from just three sample points, it is clearly an inverse relation, and close to an inverse square-root relation ($\propto \sqrt{\frac{1}{f}}$), just as (2.45) predicts. Thus, even though $\mathbf{E}$ and $\mathbf{H}$ are not orthogonal in the transverse plane for the near-field, simulations indicate that the superposition of plane waves in the angular spectrum expansion of (2.9) approximately obeys the skin depth relation. The simple concept of skin depth in this section was derived for normally incident plane waves whereas most of the waves in the angular spectrum expansion are obliquely incident. Even so, it is a useful concept for understanding trends in the interaction of a near-field probe tip with a conducting sample.

Equation (2.45) is plotted in Fig. 2.17 for various metals. From Fig. 2.16 and 2.17 we conclude that NFMM is capable of penetrating into materials, however highly conductive samples are essentially Faraday cages and prevent useful field penetration. Poor conductors, dielectrics, and metallic thin-films are much more amenable to field penetration.

We can also use HFSS to examine the size of a pixel for any given sample and probe tip. A pixel is an abstract concept, but in general it describes the localization of the field from the tip

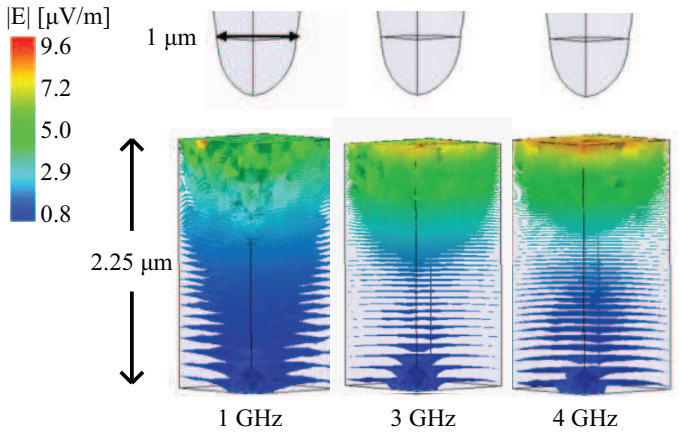


Figure 2.16: A full-wave simulation using HFSS shows the concentration of the electromagnetic field form a tip operating at 1 and 4 GHz.

on the (x, y) surface of the DUT. The field actually spreads to infinity so we define a cutoff field value of perhaps -3 dB or -10 dB. Figure 2.18 shows a 15 μm tip over 0.5 μm thick metal on Si. The -3 dB pixel diameter is 8 μm and the -10 dB pixel diameter is 19 μm, both near the physical tip diameter of 15 μm. The tip is simulated very close to the surface (inset) to eliminate pixel degradation effects due to scan height.

While HFSS is a valuable simulation tool, FEMM is better suited for rapid analysis of parameterized tip geometries including imperfections at the tip apex, as well as contributions

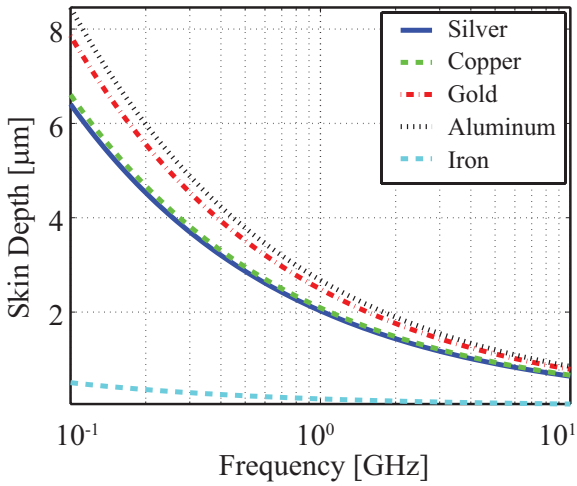


Figure 2.17: Skin depth is a function of conductivity, permeability and frequency. Most materials commonly used in this work are non-magnetic and thus $\mu_r = 1$. In general, skin depth decreases as frequency increases, and most conductors are similar. Therefore, in order to penetrate through a thin layer of metal we require low frequency electromagnetic waves.

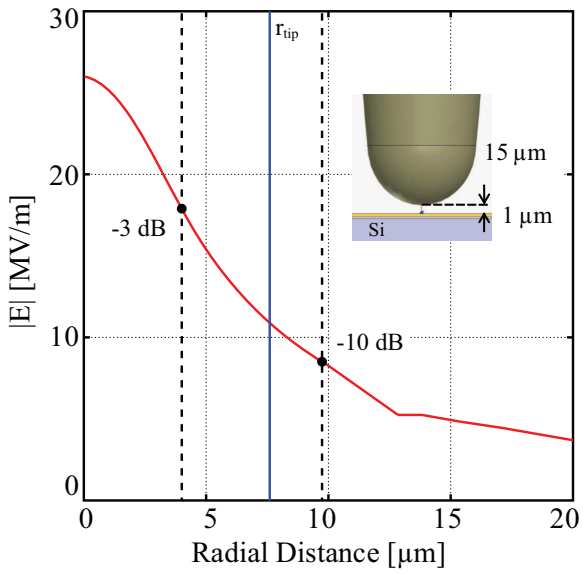


Figure 2.18: The electric field is extracted from HFSS simulations of a 15 μm diameter probe tip over a 0.5 μm conductive layer on a silicon substrate. Pixel size is defined as a circle in which the electric field decays to some threshold below the max value in the center. A 3 dB pixel threshold results in 8 μm diameter pixels while a 10 μm threshold results in a 19 μm wide pixel.

from the conical taper, and the cylindrical shaft. Through automation these rapidly converging simulations can be used as part of an error function in a least-square fitting procedure. Figure 2.19 shows the electrostatic potential around a probe model in FEMM which includes all significant elements of the tip geometry. From the field results we can calculate the total charge on the tip electrode

$$Q_{\text{enc}} = \oint_S \mathbf{D} \cdot \mathrm{d}\mathbf{S}, \tag{2.46}$$

where S is shown in Fig. 2.19. This calculation can be performed in the field calculator in HFSS by defining an S slightly larger than the tip and then calculating (2.46), while FEMM calculates it automatically. It is important that S completely surround the full tip electrode. The potential between the ground plane and the tip electrode is

$$V = -\int_0^h \mathbf{E} \cdot \mathrm{d}\mathbf{l}, \tag{2.47}$$

where $\mathrm{d}\mathbf{l}$ is shown in Fig. 2.19. This too can be performed with the field calculator in HFSS or assumed from boundary conditions in FEMM. From (2.46) and (2.47), $C_{\text{tip}} = Q_{\text{enc}}/V$. Shaft length l_{shaft}, shaft diameter d_{shaft}, taper angle θ, diameter of the tip d_{tip}, and size of the

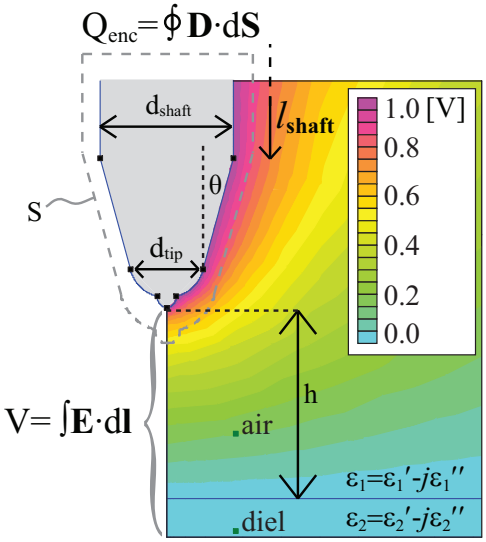


Figure 2.19: Complex probe tip models are rapidly simulated in a quasi-static field simulation to calculate capacitance as a function of many parameters for fitting in MATLAB.

hemispherical apex imperfection can be varied rapidly to determine their effect on C_{tip}.

In both the analytical methods and simulations of section 2.6 we idealized the actual configuration by assuming that the metal plane beneath the DUT is RF ground. In fact, there is some length of line between the DUT platform and RF ground, therefore there is an impedance to ground. For this reason some field lines will actually return to ground on the backside of the microwave probe circuit. This means that some of the capacitance is not a function of the DUT so we will expect a stray capacitance C_{stray}, in addition to the free space capacitance C_{∞} already discussed in section 2.6.1, which is not a function of the probe tip. If the tip is very close to the DUT the tip-DUT capacitance will dominate the parasitic capacitance, however if the tip is far from DUT the saturation capacitance $C_{\text{par}} = C_{\infty} + C_{\text{stray}}$ will dominate. NFMM measurements require the tip-DUT capacitance dominate C_{stray} so we must scan very close to the sample and minimize the stray and free-space capacitance. We revisit this concept in section 4.1.2.

2.7 Summary

This chapter overviews the concept of sub-wavelength resolution imaging and introduces the main concepts of the NFMM measurement method. The probe circuit is introduced, along with several modeling approaches including the development of an equivalent circuit. Results from this chapter were published in [21].

Chapter 3

Lock-in Vector Near-field Probe

la carrière ouverte aux talents–the tools to him that can handle them.

—Thomas Carlyle on Napoleon, in his *Mirabeau* essay, 1837.

Contents

3.1 Overview

In the previous section we discussed the principles of NFMM and the design of the near-field probe tip and microwave probe circuit. In this section we discuss the instrumentation referred to as the Lock-in Vector Near-field Probe (LVNP) including aspects of impedance measurements, the noise environment and noise mitigation techniques. We finish with several measurements to characterize the performance of the LVNP.

3.2 Lock-in Vector Near-field Probe

It is tempting to employ a network analyzer to measure the response of the probe circuit discussed in chapter 2 since the NFMM signal is simply a changing microwave impedance. However, we must measure small changes in the resonant response of the probe circuit in the presence of an otherwise large background signal which requires high dynamic range[1] . A typical NFMM measurement may take between a few minutes and a few hours, so it additionally requires extremely good long-term stability. A modern network analyzer with a low IF bandwidth and averaging is able to achieve high dynamic range, however, each independent measurement requires a relatively long period of time to complete so the drift ($1/f$ noise) of the instrument will prevent useful comparison of measurements over an entire scan area.

3.2.1 Instrumentation & Measurement Topology

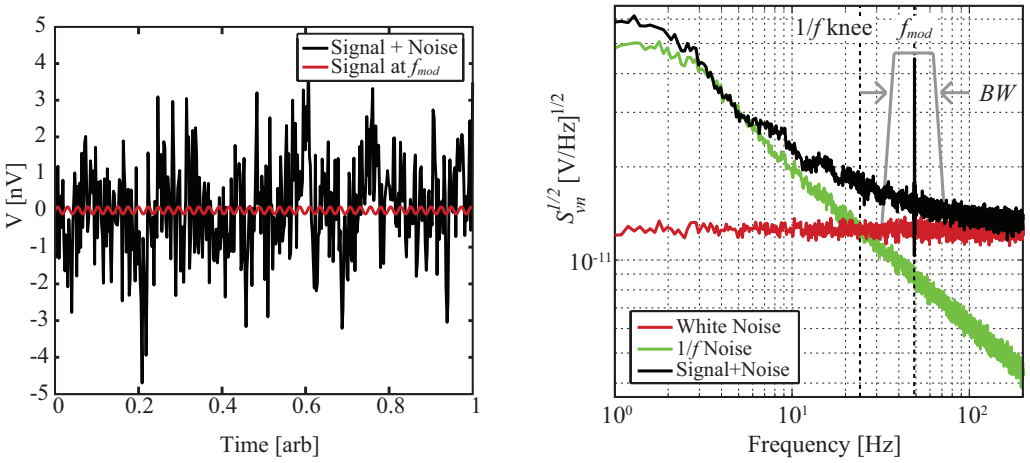
In appendix A (see Tab. A.1) we show that the variance of a measurement in the presence of $1/f$ noise for a phase sensitive detector (PSD) is non-divergent, implying that long-term stability can be achieved. Phase sensitive detection is essentially measuring a low frequency or direct

[1] Even though the probe is impedance matched to reduce background signal and thus decrease the required dynamic range, on some level the signal of interest is small with respect to the nominal reflection and thus requires high dynamic range.

current (DC) signal (e.g., Γ_{probe} at f_{op}) by modulating it beyond the $1/f$–knee (f_c) of the system, then filtering for an arbitrarily long period of time, τ_{int}. In this way the required dynamic range can be achieved with integration, but long-term stability is preserved because the measurement occurs completely beyond f_c. Though the actual phase sensitive detection (PSD) mechanism is accomplished through homodyne detection of the modulated signal with a reference signal and a low-pass filter, it can be thought of as a narrow-band (BW $\propto \frac{1}{\tau_{\text{int}}}$) band-pass filter centered at the modulation frequency f_{mod}.

Figure 3.1 shows the simulated time domain and frequency domain of a small signal of interest that has been intentionally modulated at $f_{\text{mod}} = 50\,\text{Hz}$, buried in white and $1/f$ noise. In the time domain (Fig. 3.1a) the sinusoidal signal (red) is totally overwhelmed by white and $1/f$ noise (black trace). However, in the frequency domain (Fig. 3.1b) the small harmonic signal is distinct from the broadband white noise, and the low-frequency $1/f$ noise. In fact, an arbitrarily high SNR can be achieved while still preserving long-term stability by applying PSD which is equivalent to applying a band-pass filter of $BW \propto 1/\tau_{\text{int}}$ centered at f_{mod}. Notice, however, that if $f_{\text{mod}} \leq 1\,\text{Hz}$, the noise spectrum would overwhelm the signal and no amount of integration could extract the signal from the noise. This is the case for the NFMM signal of interest without modulation. Therefore, the LVNP must explicitly modulate the probe reflection in order to employ PSD.

The LVNP is capable of performing high dynamic range one-port network analyzer measurements while also achieving the necessary stability. A simplified schematic of the LVNP is shown in Fig. 3.2. The probe circuit discussed in section 2.5 is shown to the right of the vertical dashed line, while the LVNP is to the left. The LVNP is a one-port reflectometer with modulation and phase sensitive detection. Referring to Fig. 3.2, the power from an RF local oscillator (LO) is split: one part is reflected off the probe circuit (through a circulator) and mixed in phase quadrature with the other part. For constant LO power the voltages at the I and Q ports are



(a) Small sinusoid buried in noise (time domain)

(b) Corresponding frequency domain spectrum with a spike at the modulation frequency f_{mod}.

Figure 3.1: (a) A 50 Hz sinusoid (red) is buried in white and $1/f$ noise (blue) in the time domain. (b) In the frequency domain it is clear that the modulated sinusoid could be extracted from the broadband white noise and low-frequency $1/f$ noise with a phase sensitive detector, or lock-in amplifier. Post-detection integration of τ_{int} is equivalent to a bandpass filter centered around f_{mod} of $BW \propto 1/\tau_{\mathrm{int}}$.

DC and proportional to the magnitude and phase of the reflection coefficient:

$$V_I \propto |\Gamma_{\mathrm{probe}}| \cos\left(\angle\Gamma_{\mathrm{probe}}\right) \tag{3.1}$$

$$V_Q \propto |\Gamma_{\mathrm{probe}}| \sin\left(\angle\Gamma_{\mathrm{probe}}\right), \tag{3.2}$$

such that two measured quantities, V_I and V_Q, uniquely determine the complex probe reflection coefficient Γ_{probe} (also referred to as S_{11} throughout this work). This is the one-port reflectometer, and by employing arbitrarily long integration times, $V_{I,Q}$ can be measured with the required dynamic range.

By adding a modulation source at f_{mod} and two Stanford Research lock-in amplifiers V_I and V_Q are shifted from DC to f_{mod} and thus can be extracted from seemingly overwhelming noise. Fig. 3.2 shows two modulation schemes: (1) LO modulation, in which the RF LO is amplitude modulated at f_{mod}, and (2) piezo modulation, in which the height of a piezoelectric motion stage is modulated such that the scan height h is modulated at f_{mod}. Both methods effectively translate the otherwise DC I and Q signals to $f_{\mathrm{mod}} > f_c$ allowing for long-term stability. The effects of each are discussed further in section 3.3.1. Figure 3.3 shows the de-modulated I and Q

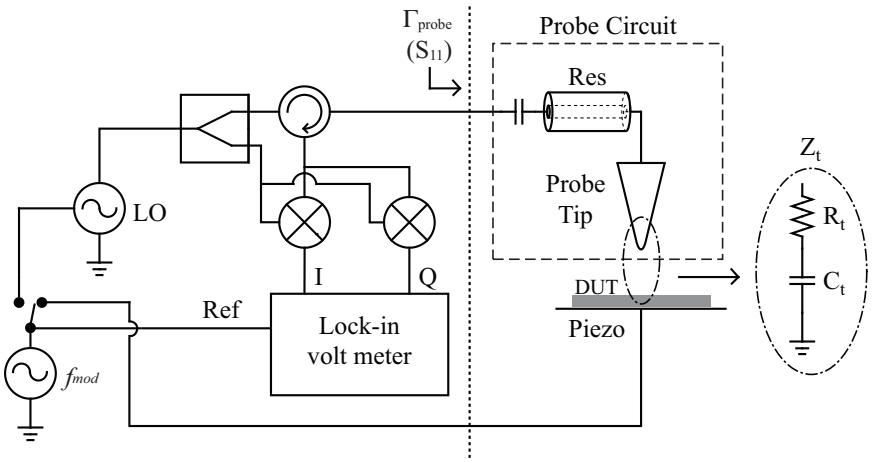
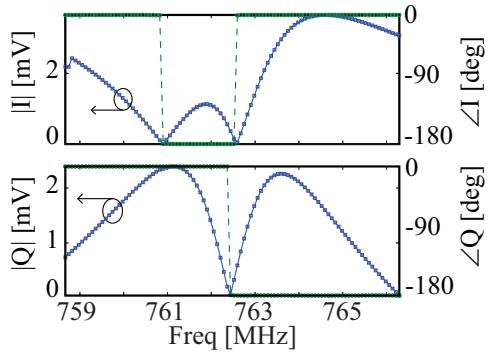
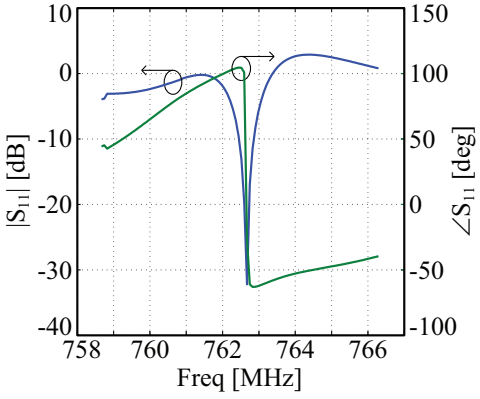


Figure 3.2: A microwave excitation (LO) is divided such that part of the power is reflected off the probe circuit (dashed outline) and mixed in quadrature with the other part of the power. For a CW excitation, at a single location on the DUT the resulting I and Q are DC signals proportional to Γ_{probe}. When the LO or the Piezo is modulated at f_{mod} then I and Q are AC signals proportional to Γ_{probe} and thus the LVNP can employ phase sensitive detection.

signals (a) and the corresponding S_{11} (b) from a frequency sweep of the probe circuit with the LVNP. Due to poor calibration in Fig. 3.3 $|S_{11}| > 0\,\text{dB}$ above the resonance.



(a) Demodulated I and Q outputs from the LVNP.



(b) S_{11} calculated from I and Q.

Figure 3.3: (a) The I and Q outputs from a frequency sweep are the in-phase and quadrature components of the reflection coefficient, and (b) the corresponding S-parameters.

3.2.2 Measurement Principle

We demonstrate the measurement principle of the LVNP through a typical measurement in which the probe circuit and probe tip are held fixed while the DUT is laterally scanned along the x-dimension. The DUT is a metallic test structure with a well defined step transition as shown

schematically at the top of Fig. 3.4. We discuss two points along the scan trajectory, x_1 close to the DUT step transition, and x_2 further from the step. The DUT is scanned to each location and held for a fixed time while the piezo motion stage is vertically modulated at f_{mod}. Therefore the separation distance between the DUT and the probe tip is modulated between a height h_0 and $h_0 + \delta h_{\text{piezo}}$. We refer to the two states as the "nominal" and "modulated" states respectively, and assume the nominal states at x_1 and x_2 are far enough from the DUT that they have the same coupling to the DUT (Measurements confirm this to be a good approximation). However, the coupling of the modulated states differs between x_1 and x_2.

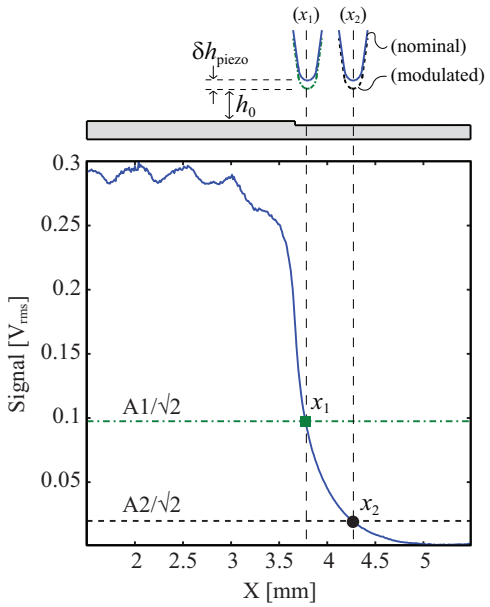
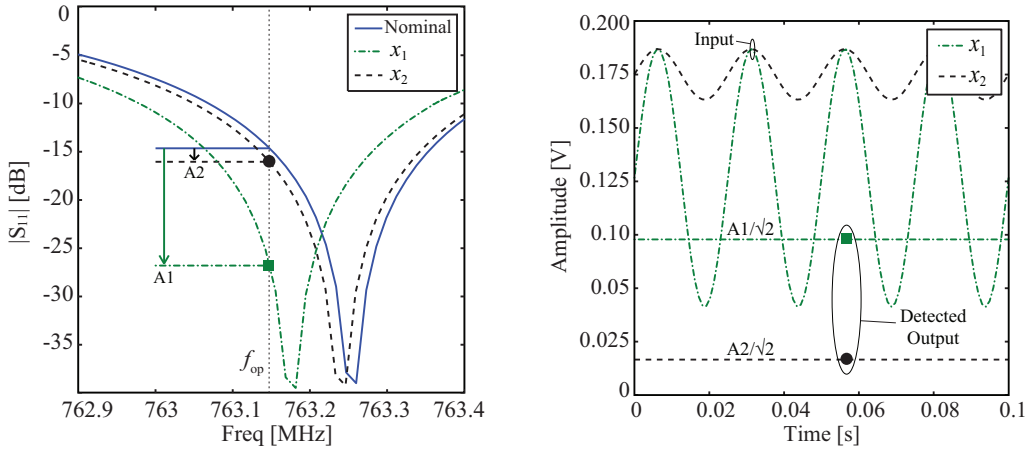


Figure 3.4: A lateral scan across a metallic step transition is shown. Here the LVNP employs piezo modulation in which the nominal scan height $h_0 \approx 1\,\mu$m is modulated by an amount $\delta h_{\text{piezo}} \approx 100\,$nm between a "nominal" and "modulated" state. The measurement (solid trace) is the result of 501 data points, two of which will be focused on for a description of the measurement: x_1 (square marker) and x_2 (circle marker).

Figure 3.5a shows the frequency response of the loaded probe circuit for the nominal state (same at x_1 and x_2), the modulated state at x_1, and the modulated state at x_2. As the tip nears the sample increased capacitive loading effectively lengthens the resonator shifting the resonant frequency down. When the LVNP is operated at a single frequency f_{op}, the input to the PSD (Fig. 3.5b) is a sinusoid at f_{mod} varying between the nominal and modulated reflected voltages. The output is the DC rms equivalent of the amplitude of the modulated signal and is the recorded

signal at x_1 and x_2. Figure 3.4 and Fig. 3.5a show measured data, but Fig. 3.5b is an illustration because in reality the sinusoidal inputs to the phase sensitive detector are buried in noise and cannot usually be measured, hence the need for phase sensitive detection.



(a) $|S_{11}|$ of probe circuit.

(b) Input and output of phase sensitive detector (lock-in).

Figure 3.5: (a) At x_1 in the lateral scan the probe circuit response ($|S_{11}|$) is modulated between the nominal response (solid trace) and the response due to tip-DUT coupling at x_1 (dashed trace). At x_2 the response is modulated between the nominal response (same as at x_1) and the response due to tip-DUT coupling at x_2 (dash-dot trace). At f_{op} the difference between nominal and x_1 is A_1 while the difference between nominal and x_2 is A_2. (b) The inputs to the phase sensitive detector (lock-in amplifier) are sinusoids of peak value A_1 and A_2 at the modulation frequency f_{mod} (frequency of piezo height variation, *not* the RF frequency f_{op}). The DC offset represents the background signal which is filtered out by the lock-in. The outputs of the phase sensitive detector are DC voltages of RMS $A_1/\sqrt{2}$ and $A_2/\sqrt{2}$. These two data-points, in addition to all other x-values in the lateral scan make up the solid trace in Fig. 3.4.

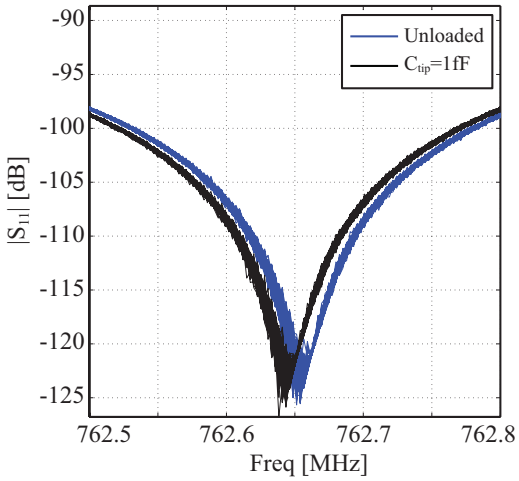
3.2.3 Operating Point

The LVNP can be operated in either a frequency sweep mode or a rapid single frequency mode. That is, at every (x, y) location in a scan S_{11} can be recorded over a range of frequencies around f_0, or at a single f_{op} frequency near f_0. When developing models it is useful to have as much information as possible so a full frequency sweep around the resonance of the probe circuit is employed. However, when scanning a large area it is often impractical to measure more than a single frequency point at each (x, y) location in the scan. In addition, drift in the LVNP will have less effect on a fast measurement. The question then becomes how to select the best operating point. In the literature it is often suggested that the operating point be at a fixed frequency

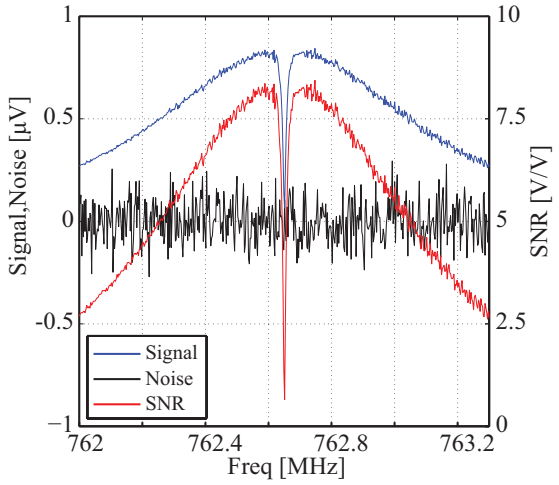
point slightly lower than the resonance frequency [51], however we find this is not generally true. Optimal operating frequency is a function of tip impedance.

We demonstrate by simulating a frequency sweep of the probe circuit from Fig. 2.7a and adding a theoretical white noise process of amplitude approximately equal to that in a typical measurement ($\sigma_n = 1\mu$V). An unloaded probe response is shown together with a 1 fF loaded probe response in Fig. 3.6a. The output signal from the PSD due to this change in tip capacitance is the difference between the two sweeps at each frequency point, as plotted in Fig. 3.6b. Because the noise is white, the signal-to-noise ratio (SNR) follows the same functional trend as the signal trace. For the traces shown in Fig. 3.6b the optimal operating point f_{op} is slightly below f_0. By way of example, we measured a large probe tip ($d_{\text{tip}} \approx 200\,\mu$m) scanned over a square metal patch surrounded by dielectric at several operating frequencies. Figure 3.6c shows the resulting maximum signal difference (metal vs. dielectric) as a function of offset from resonance, Δf_{op}. Despite the coarse sampling, the general trend is the same as the expected double hump from Fig. 3.6b. However, if the tip were a different size or scanned at a different height then $\mathrm{d}C_{\text{tip}}$ would change and the loaded frequency sweep would shift resulting in a maximum SNR at a different frequency. In general the optimal operating frequency is a function of tip impedance, but since scan height should be set according to the tip diameter ($h \approx d_{\text{tip}}$), then the optimal operating frequency can be set for each tip diameter as shown in Fig. 3.7 where we have simulated the nominal resonance due to probe tips of various diameter and then plotted the maximum signal as a function of offset from the resonance.

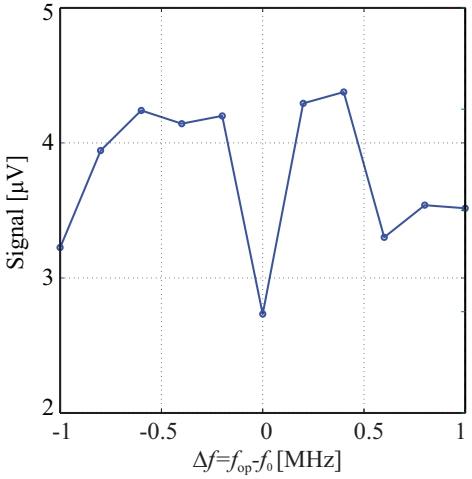
In order to correctly set f_{op} for every measurement we perform a frequency sweep of the probe circuit at the beginning of each measurement to determine the actual resonant frequency, f_0, and then set $f_{\text{op}} = f_0 - \Delta f_{\text{op}}$ for the particular tip diameter from Fig. 3.7. Notice that for small tip diameters ($< 10\,\mu$m) the optimal offset frequency stabilizes at approximately 55 kHz. Though f_0 is almost the same every time small changes in cable position, component temperature, connector torque, etc. slightly shift the resonance. The expected signals in an NFMM measurement are so small that we need to know the measured resonance precisely in order to operate at the optimal



(a) Nominal resonance plus white noise.



(b) SNR.



(c) Signal and noise measured.

Figure 3.6: (a) Probe circuit resonance with a theoretical white noise process added. (b) Theoretical SNR assuming a white noise limiting process. (c) Measured signal and noise as a function of offset frequency from probe circuit resonance f_0.

RF frequency. Figure 3.8 shows the procedure for finding the resonant frequency begining with a coarse frequency sweep and then continuously zooming in on the resonance until the measured f_0 does not change by more than 10 kHz, or 0.001% of f_0. Then the LVNP is set to operate at f_{op} and the measurement proceeds.

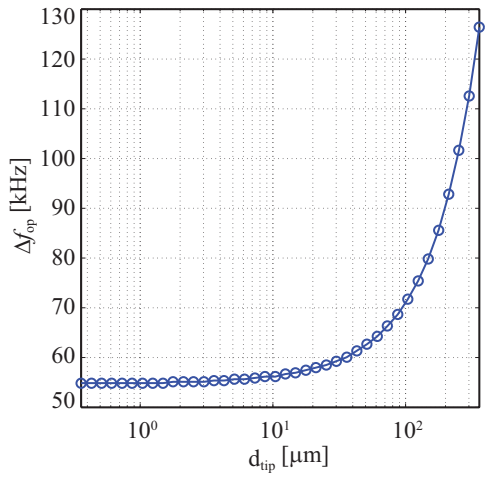


Figure 3.7: The optimal offset frequency, where signal is greatest, is a function of tip-diameter only so long as the tip is scanned at $h = d_{\text{tip}}$.

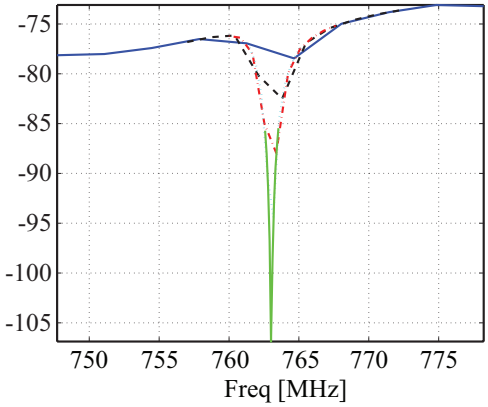


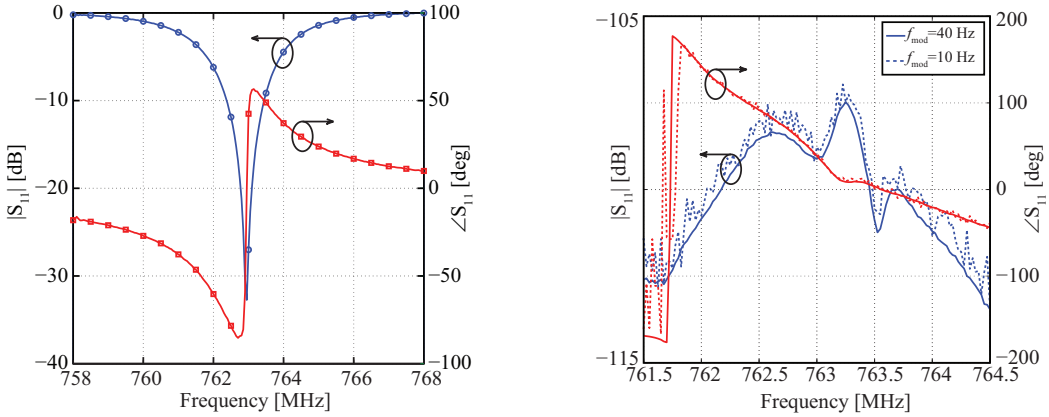
Figure 3.8: The precise resonant frequency is measured before each scan is conducted by a series of successively finer frequency sweeps about the resonant frequency.

3.3 System Characterization

The previous section described the LVNP instrument for NFMM measurements. The LVNP is capable of arbitrarily long integration times for required SNR, and arbitrary modulation frequencies with one of two modulation mechanisms. In this section we examine the effects of modulation mechanism (LO or piezo), modulation frequency, general noise and stability performance, vibration and positioning artifacts, and system calibration.

3.3.1 Modulation Effects

To compare the effects of modulation scheme—LO or piezo—we examine frequency sweeps of the probe circuit. Figure 3.9 shows frequency sweeps with LO modulation at 40 Hz (Fig. 3.9a) and piezo modulation at 10 and 40 Hz (Fig. 3.9b). They differ in three significant ways: modulation *mechanism*, *calibration*, and *signal-to-noise ratio (SNR)*.



(a) LO modulation.

(b) Piezo modulation.

Figure 3.9: The LVNP can modulate the amplitude of the LO or the height of the piezoelectric motion stage. (a) shows a calibrated LO modulation frequency sweep, while the piezo modulated frequency sweep (b) cannot be calibrated. (b) also shows that a frequency sweep with 40 Hz piezo modulation is significantly less noisy than one with 10 Hz, demonstrating the benefits of variable modulation frequency.

While the piezo modulation mechanism is height variation of the tip-DUT separation distance, the LO modulation mechanism is amplitude modulation of the LO. That is, the tip-DUT separation distance remains constant, but the RF voltage at the probe is modulated at f_{mod}. The output of the PSD is the difference between the nominal and modulated states. For LO modulation this is the difference between a resonance and a highly attenuated version of the same resonance. Therefore, the plot in Fig. 3.9a has the same resonant shape as would be measured by a VNA. The output of the PSD for a piezo modulated frequency sweep is the difference between a resonance and a slightly shifted resonance which results in a double-hump as in Fig. 3.9b. For reasons discussed in section 3.3.5, LO modulated measurements can be calibrated while piezo modulated measurements cannot. Therefore the measurement in Fig. 3.9a is calibrated while a

systematic error appears at 763.5 MHZ in Fig. 3.9b.

Although piezo modulation does not support calibration, it does provide higher SNR than LO modulation for a comparable integration time. This is because in any NFMM measurement the actual signal of interest is the comparatively small perturbation of a larger nominal resonance. In the case of piezo modulation the PSD differences two shifted resonances which effectively subtracts out the large background signal and leaves only the signal of interest, while for LO modulation the PSD differences a resonance with something close to zero preserving the majority of the background signal.

While LO and piezo modulation differ in many respects, both can be operated over a continuous range of modulation frequencies which is beneficial for noise reduction and interference immunity. The piezo can be modulated at ($f_{\text{mod-piezo}} < 100\,\text{Hz}$), limited by mechanical resonances in the scanning hardware. The LO can be modulated at ($20\,\text{Hz} < f_{\text{mod-LO}} < 100\,\text{kHz}$), limited by the modulation hardware of the specific LO used in this work. As mentioned in section 3.2.1, PSD can be thought of as a narrow band band-pass filter around the modulation frequency. Random noise within the equivalent noise bandwidth (ENBW)[52] of the filter will contribute to overall measurement noise and adjacent interference sources not attenuated sufficiently by the filter will couple into the measurement. Judicious selection of f_{mod} can dramatically reduce the integration time required to achieve the desired SNR.

Figure 3.9b shows piezo modulated frequency sweeps at $f_{\text{mod}} = 10\,\text{Hz}$ and $f_{\text{mod}} = 40\,\text{Hz}$. Even though both frequency sweeps employ a 300 ms integration time, the 10 Hz sweep includes interference and is generally more noisy than the 40 Hz sweep. Referring to the system noise spectrum in Fig. 3.10, this is because 10 Hz is far below the $1/f$ knee, f_c, while 40 Hz is much closer to f_c in a region with a lower noise floor (discussed further in Section 3.3.2). It would be possible to increase τ_{int} thereby narrowing the BW around 10 Hz to eliminate the interference but depending on how strong and close the interferer is this may require a prohibitively long integration time. With a properly selected modulation frequency a given measurement will require less total time.

It has been mentioned that integration time τ_{int} is inversely related to BW. Here we show this relationship explicitly through the concept of equivalent noise bandwidth (ENBW) for a low-pass filter (LPF) and an ideal averaging filter (IAF) [53] (see section A.3.1), in the presence of white noise[2] . The output variance of a measurement is, from (A.2):

$$\sigma^2_{v\text{rms}} = \int_0^\infty \left(S^{1/2}_{v\text{nBB}}\right)^2 |G(j2\pi f)|^2 \mathrm{d}f, \tag{3.3}$$

where $S^{1/2}_{v\text{nBB}} = \sqrt{4kTR}$ is the single-sideband white thermal noise voltage spectral density with units of $[\text{V}/\sqrt{\text{Hz}}]$, and $|G|$ is the transfer function. Both terms in the integrand are modulus squared because we operate on noise powers. The transfer functions for a LPF and an IAF are

$$G(j\omega) = \begin{cases} \frac{1}{1+j\omega\tau} & \text{LPF} \\ \exp(-j\omega\tau/2)\frac{\sin(\omega\tau/2)}{\omega\tau/2} & \text{IAF} \end{cases} \tag{3.4}$$

where the traditional $\omega_0 = 1/\tau$. The complex modulus of the LPF transfer function is

$$|G(j\omega)|_{\text{LPF}} = \frac{1}{\sqrt{(1)^2 + (\omega/\omega_0)^2}}, \tag{3.5}$$

and since $S_{v\text{nBB}}$ is constant over frequency (i.e., white noise)

$$\begin{aligned} \sigma^2_{v\text{rms}} &= S_{v\text{nBB}} \int_0^\infty \frac{1}{1 + \left(\frac{f}{f_0}\right)^2} \mathrm{d}f, & (3.6)\\ &= S_{v\text{nBB}} \left[f_0^2 \left(\frac{1}{f_0} \tan^{-1}\left(\frac{f}{f_0}\right)\right)\right]_0^\infty & (3.7)\\ &= S_{v\text{nBB}} \left[f_0 \left(\tan^{-1}(\infty) - \tan^{-1}(0)\right)\right]_0^\infty & (3.8)\\ &= S_{v\text{nBB}} f_0 \frac{\pi}{2} & (3.9) \end{aligned}$$

and

$$v_{\text{rms}} = S^{1/2}_{v\text{nBB}} \sqrt{f_0 \frac{\pi}{2}}. \tag{3.10}$$

The equivalent noise bandwidth for thermal noise is defined as

$$ENBW = \sqrt{f_0 \frac{\pi}{2}} = \sqrt{\frac{1}{4\tau}}. \tag{3.11}$$

[2] We assume the measurement employs PSD such that $1/f$ noise is negligible.

That is, ENBW is the bandwidth of an effective brick-wall filter on the noise. Following the same procedure for the IAF

$$ENBW = \begin{cases} \frac{1}{4\tau} & \text{LPF} \\ \frac{1}{2\tau} & \text{IAF.} \end{cases} \tag{3.12}$$

Therefore, the equivalent bandwidth for calculating the effect of white noise is inversely proportional to integration time τ.

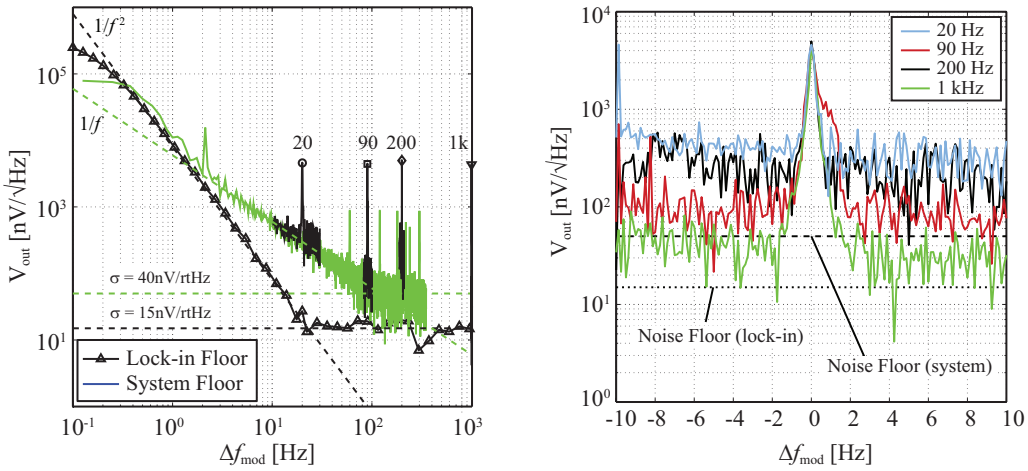
3.3.2 Noise and Stability

NFMM measurements can take several minutes to several hours to complete depending on the desired SNR, the total area scanned, and the required spatial resolution. For this reason, much effort has been dedicated to developing a system architecture (the LVNP discussed in section 3.2.1) that provides low noise and long-term stability. In the previous section we made reference to the noise spectrum several times but did not formally examine it. In this section we measure the noise spectrum of the LVNP and determine the noise-limiting elements, then measure long-term stability.

Figure 3.10a shows the noise floor of the lock-in amplifier and the overall LVNP system noise floor at $15\text{nV}/\sqrt{\text{Hz}}$ and $40\text{nV}/\sqrt{\text{Hz}}$ respectively. The noise spectra were measured by sweeping the reference signal of the lock-in amplifier between 0.1 and 350 Hz and recording the power in each frequency bin, during which the experiment was not modulated[3] . The system noise floor has a $1/f$ slope as shown in dashed green. $1/f$ noise has a $1/f$ power spectrum implying a $1/\sqrt{f}$ voltage spectrum. However, in this case the voltage is proportional to input power so the slope is $1/f$. Below 1 Hz the $1/f^2$ noise of the lock-in dominates the system noise. In order to determine which component of the LVNP limits the noise floor we sequentially short the input to various elements and perform a spectrum measurement. The dashed trace labeled "Lock-in Floor" shows the measured noise spectrum with the inputs to the lock-in amplifier shorted and defines the lowest noise floor of the system with a $1/f$ knee at $f_c = 20$ Hz. The solid trace labeled "System Floor" shows the noise spectrum with the inputs to the I- and Q-mixers shorted. The mixers

[3] This is the same procedure as an RF spectrum analyzer.

raise the white noise floor and degrade f_c from 20 Hz to approximately 100 Hz. Sequentially shorting all the way to the RF input of the LVNP did not increase the noise floor any more, therefore the IQ-demodulator is the noise limiting element. This is not surprising because the commercial IQ-demodulator includes an active LO amplifier which has a non-zero noise figure and significant $1/f$ noise. The noise floor could be reduced by placing an LNA before the LO port or by making a fully passive IQ-demodulator. We revisit this option in section 5.2.1.



(a) Noise spectra with various LO modulation frequencies.

(b) LO modulation frequency noise floors superimposed.

Figure 3.10: (a) The LVNP is LO modulated at 20 Hz, 90 Hz, 200 Hz, and 1 kHz and the resulting noise spectra is recorded. The system $1/f$ knee is approximately 100 Hz and in general the LO modulated noise floor is very close to the unmodulated noise floor. (b) The noise floor within a 20 Hz bandwidth around the LO modulation frequency shows the noise floor at the various LO modulation frequencies overlaid. At 20 Hz modulation the noise floor has a $1/f$ slope while at 90 Hz the floor is nearly flat as it is very close to the $1/f$ knee. At 1 kHz the noise floor is lowest, but at 200 Hz the noise floor is anomalously high, probably due to nearby strong interferes from power line harmonics at 180 and 240 Hz.

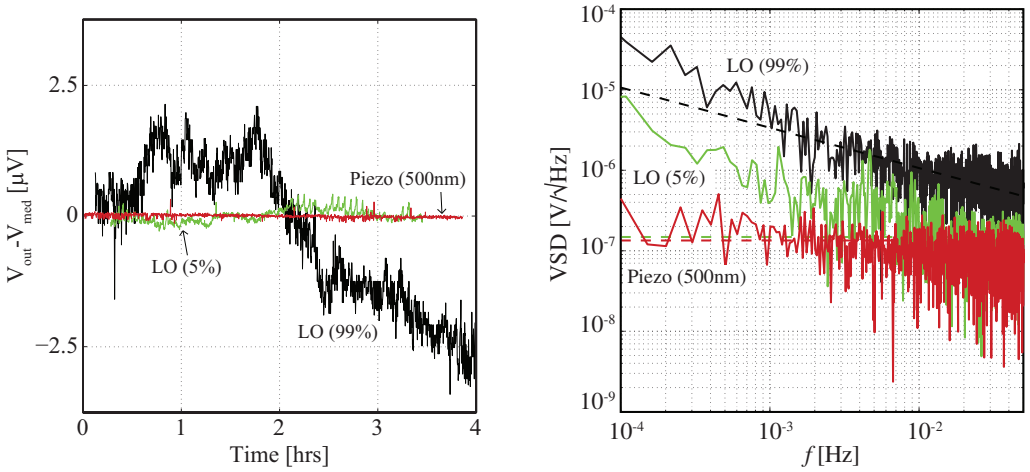
Fig. 3.10b shows the superimposed noise floors at various LO modulation frequencies: 20 Hz, 90 Hz, 200 Hz, and 1 kHz, over a 20 Hz bandwidth. It is clear that choice of modulation frequency effects overall noise floor (assuming most operational integration times yield measurement bandwidths less than the ±20 Hz shown here). Of particular note is the $1/f$ slope for the 20 Hz trace indicating the modulation frequency is not beyond the $1/f$ knee. However, at 90 Hz, the $1/f$ slope is almost imperceptible because it is so close to the $1/f$ knee. The 1 kHz modulation is completely in the white noise regime and has the lowest noise floor overall. Modulation at 200 Hz

gives rise to an anomalously high noise floor which we attribute to its close proximity to very strong interferers at harmonics of the line frequency 60 Hz.

Next we would like to characterize long-term stability by recording a nominally constant output over time and observing fluctuations from nominal. The time series of such an output is plotted in Fig. 3.11a where the median value has been subtracted ($V_{\text{out}} - V_{\text{med}}$), and Fig. 3.11b shows the corresponding FFTs. We compare two LO modulation depths: 5%, and 99%, and piezo modulation of 500 nm. Piezo modulation exhibits extremely good stability over four hours and is nearly white in the frequency domain down to 100 μHz. As LO modulation depth increases, stability degrades and $1/f$ noise becomes dominant. However, even in the case of 99% LO modulation depth the $1/f$ knee is $f_c = 0.01$ Hz, which is four orders of magnitude below the noise spectra shown in Fig. 3.10a. It is important to understand that the noise spectra in Fig. 3.11b were measured with modulation and phase sensitive detection enabled while Fig. 3.10a had phase sensitive detection disabled. In other words the low frequency noise in Fig. 3.10a is the $1/f$ noise of the electronics without PSD. The dramatic reduction of f_c when modulated at f_{mod} is the direct result of the PSD built into the LVNP, without which, long duration NFMM measurements would not be possible. To achieve the best stability for a large scan with fine spatial resolution and good SNR, shallow LO modulation or piezo modulation should be used.

3.3.3 Vibration & Positioning

In order to examine vibration noise we again conduct noise spectra measurements as in section 3.3.2, but this time with the probe tip strongly coupled to a DUT (i.e., $h \approx d_{\text{tip}}$), and the entire scanning hardware, DUT, and probe circuit set on a laboratory bench (i.e., no vibration isolation). Figure 3.12 again shows the lock-in noise floor, the electronic system noise floor, but now we include the noise floor where vibration noise dominates electronic noise (labeled "Vibration"). The vibration noise is a complicated spectrum of spikes across the entire measured range. One would be hard pressed to find a stable frequency to modulate the LVNP without a strong interferer nearby. For this reason, in order to achieve repeatable NFMM measurements an



(a) Drift (time domain)

(b) $1/f$ noise (frequency domain)

Figure 3.11: Drift in the time domain is equivalent to $1/f$ noise in the frequency domain. (a) Time traces of a constant output signal over four hours, and the corresponding FFTs, show that piezo modulation is extremely stable while increasing LO modulation depth results in decreased long-term stability. This is a result of the PSD differencing the signal in the high state with a low state that has dropped below the white noise floor.

acoustic and vibration isolation chamber is required [54], as shown in Fig. 3.13.

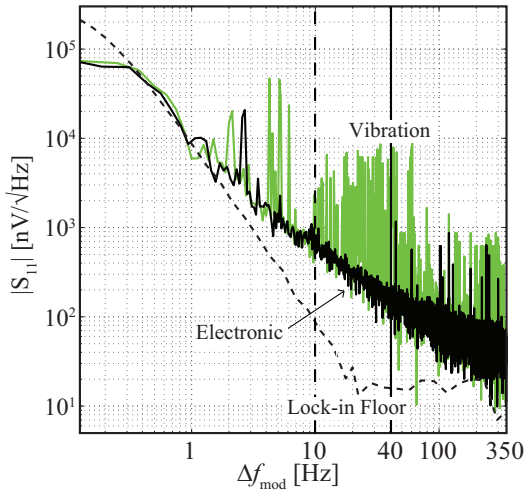


Figure 3.12: The noise spectrum for the un-modulated LVNP, limited by the I- and Q-mixers (solid trace labeled "Electronic") shows a $1/f$ region extending to approximately 200 Hz while the lock-in amplifier (dashed trace) has a $1/f$ knee at 20 Hz. When the LVNP is operated without vibration isolation (solid trace labeled "Vibration") the noise spectrum is dominated by vibration noise.

The chamber is comprised of an acoustically insulated box, with a 30 kg granite slab suspended by four bungee-cords. Acoustic foam and an enclosed box provide isolation from acoustic waves

due to conversation and turbulent air currents. The granite and bungee-cords form a mechanical single-pole low-pass filter with a cutoff frequency around 20 Hz.

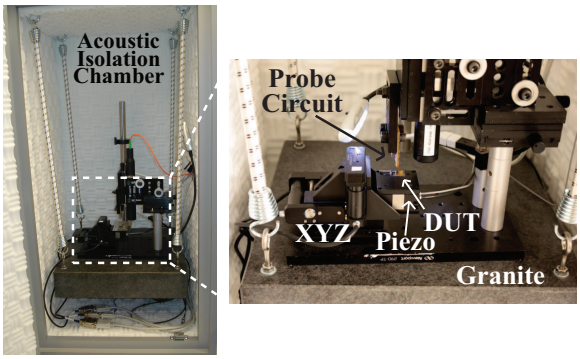


Figure 3.13: An acoustic and vibration isolation chamber houses the probe, DUT, and scanning hardware. The chamber includes acoustic isolation foam, and a mechanical low-pass filter constructed of a 30 kg granite platform suspended by four bungee-cords and yielding a cutoff frequency of approximately 20 Hz. The DUT sits on a piezoelectric z-axis motion stage which is scanned beneath the probe tip by xyz-axes mechanical motion stages.

3.3.4 Positioning Artifacts

We also need to examine errors due to piezoelectric positioning artifacts. It is well known that piezoelectric devices exhibit hysteresis and so they are often controlled with feedback. For this measurement we will use an optical profilometer to record the actual position of the piezoelectric stage compared with the control signal. Because we have an optical profilometer we also want to measure peak-to-peak vibrations directly (versus the indirect noise spectra measurement from the previous section).

Figure 3.14 shows the measured height of the DUT platform as a function of piezo motion stage control signal as the piezo is stepped from a nominal height of 0 μm to 5 μm and back to 0 μm again. The red trace is the result of stepping up, while the green trace is stepping down. The black and blue traces correspond to up and down but with a "return to zero" (RZ) protocol employed. That is, before moving to a new height the piezo is moved back to zero to erase any hysteretic memory. We note three effects in the measurement: (1) a threshold value, (2) hysteresis, and (3) vibration. The turn-on threshold is about 1.5 μm so in order to completely mitigate (1) the piezo is always operated with a slight offset to eliminate non-linearities due to

the turn-on threshold. The red and green trace shows significant hysteresis of 128 nm over a step of 5 μm. However, the blue and black trace corresponding to an RZ scan only shows 24 nm of hysteresis (the measurement noise of the profilometer is about 20 nm so we cannot measure hysteresis less than this) indicating that the RZ protocol is effective in mitigating hysteresis. Table 3.1 summarizes the hysteresis measurements.

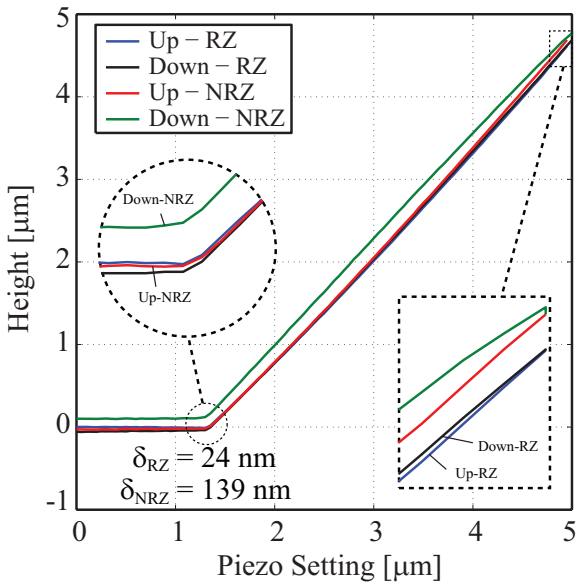


Figure 3.14: Piezoelectric hysteretic effects result in a difference of 128 nm over a 5 μm excursion with non-return-to-zero control. With return-to-zero control the difference goes down to 24 nm.

Table 3.1: Vibration and hysteresis measurements

Test	Pk-to-Pk [nm]	Hysteresis [nm]
On Table RZ	700	24
On Table NRZ	700	139
On Table (large bump)	40,000	-
On Table (small bump)	22,000	-
Vib Chamb (fast RZ)	1000	24
Vib Chamb (slow RZ)	16	24
Vib Chamb NRZ	24	139

We also examined peak-to-peak vibration under various conditions to corroborate the conclusions of Fig. 3.12. As established, NFMM measurements are not possible outside of a vibration isolation chamber. However, inside the chamber we measure peak-to-peak vibrations on the order of the accuracy of the profilometer (10-20 nm). If the piezo is modulated at too high a frequency

(beyond its mechanical resonance) the peak-to-peak vibrations increase dramatically to 1 μm.

3.3.5 Calibration

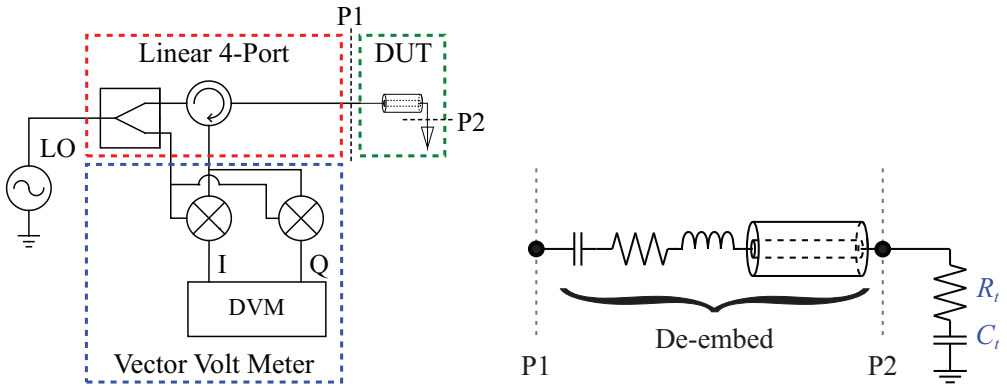
In section 2.5.2 it was assumed that the simple probe circuit model in Fig. 2.7a could accurately fit measured data. However, systematic errors throughout the LVNP prevent this unless the system is calibrated and the reference plane set to the input of the probe circuit (i.e., through the SMA connectors and onto the microwave substrate). One method of accomplishing this is to precisely characterize each component in the system with a well calibrated VNA and then de-embed their non-ideal effects in software. However, this method is susceptible to component thermal drift and imperfect connector mating. A better method is to employ a VNA-type calibration in which a number of standards are measured in place of the probe circuit. Not only does this obviate the repeated connection of every element in the system, it is also much simpler and can be performed before each measurement to better account for component drift. Once the LVNP is calibrated to the plane P1 in Fig. 3.15a, the probe circuit model from Fig. 2.7a can be de-embedded and the reference plane moved to the tip-DUT plane at P2. Only at P2 are measured impedances equal to the tip-DUT coupling impedance.

Figure 3.15a shows a block diagram of the LVNP with standard one-port reflectometer elements identified in dashed outlines. The LO serves as the excitation, the power splitter and circulator are the linear four-port, the IQ-demodulator and lock-in amplifier are the vector volt meter, and the entire probe circuit and probe tip-DUT coupling is the DUT. By fabricating custom short, open, and matched-load (SOL) impedance standards on the same substrate with the same SMA-to-planar transition as the probe circuit (shown in Fig. 3.16) we can employ the well known SOL calibration algorithm [55]. As in a network analyzer, measurement of the calibration standards defines error terms: e_{11}, $e_{21}e_{12}$, and e_{22}, which represent the linear 4-port and various scaling coefficients. We note here that it is only possible to represent a 4-port network with a 2-port error network because the reflection coefficient at two of the ports (presumably those connected to the vector volt meter) are well known and typically assumed to be well matched.

The details of this transformation are provided in [56] and [57]. The calibrated probe circuit reflection coefficient Γ_{probe} in terms of the error terms e_{mn} and the measured reflection coefficient Γ_m is:

$$\Gamma_{\text{probe}} = \frac{\Gamma_m - e_{11}}{e_{21}e_{12} - e_{11}e_{22} + \Gamma_m e_{22}}, \tag{3.13}$$

referenced to the P1 plane.



(a) One-port reflectometer calibration. (b) Probe circuit de-embedding.

Figure 3.15: (a) The LVNP is similar to a one-port reflectometer with: RF excitation (LO), linear four-port, DUT, and vector volt meter. This means the LVNP can be calibrated with an SOL calibration set like a standard VNA port. (b) The SOL calibration defines reference plane P1. The circuit model developed in Fig. 2.7a is then de-embedded moving the reference plane to the tip-DUT load at P2.

Because the PSD in the LVNP only measures the difference between the nominal and modulated states, piezo modulation cannot be calibrated with the sma short-open-load (SOL) calibration standards presented in this section. This is because when a calibration standard is connected to the LVNP and the piezo is modulated, the signal reflected from the calibration standard does not change, therefore the PSD will output only noise. A piezo modulated measurement does not interact with the calibration standards while LO modulated measurements can be calibrated because the RF excitation does interact with the calibration standards. If calibration standards are created for the DUT platform [25] then the piezo modulated LVNP could be calibrated.

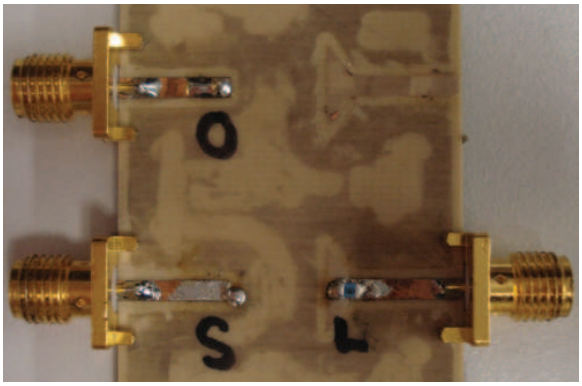


Figure 3.16: Custom short, open, and load calibration standards were fabricated on the same substrate as the microwave probe circuit with the same SMA connectors and 10 mm of microstrip transmission line. This calibration set allows the reference plane to be aligned with the critically coupled resonator and thus the probe circuit model corresponds to calibrated measurement data.

3.4 Summary

In this chapter, details of a Lock-in Vector Near-field Probe (LVNP) were presented. The instrument operation was described through a typical measurement. Both electrical (LO) and mechanical (piezo) modulation were analyzed with respect to expected output signal, noise reduction, and long-term stability. Finally, calibration of the instrument was presented. The contributions in this chapter are presented in [58, 59, 60].

Chapter 4

Measurements and Analysis

Our future discoveries must be looked for in the sixth place of decimals.

—A. A. Michelson, in *Light Waves and Their Uses*, 1903.

Contents

4.1 Measurements

With the probe circuit optimized and modeled, and noise and stability performance for the instrument characterized, and a calibration routine created, we demonstrate several LVNP measurements and perform the appropriate data analysis. Though the number of NFMM mea-

surements are limited only by the imagination, there are three general modalities discussed in this chapter. Each serves a different purpose and often requires different calibration and/or analysis methods:

(1) *Differentiation* results in either a 1D or 2D dataset which is used to classify or bin samples or regions of the same sample. This measurement differs from imaging in that it should be conducted rapidly [11] so one might only perform spot measurements.

(2) *Analysis* includes calibration of the LVNP which, in combination with various models of the probe circuit and the physical interaction of the probe tip and the sample [61], result in quantitative parameters of the tip-DUT coupling.

(3) *Imaging* results in a 2D image representing variation of the measured parameter over the scan area. It is a qualitative measurement of the composite interaction of all of the complex DUT parameters (σ, ϵ, μ) with the near-field probe [51].

A combination of the *analysis* and *imaging* modalities is referred to as quantitative imaging. This is at the forefront of the field [25] and will be viewed as the eventual goal, but has so far only been demonstrated for special cases (e.g., analytical solutions at zero scan height or curve fitting to known samples [26]). Calibration and processing differs for each modality and is discussed as appropriate throughout.

For the measurements that follow, custom MATLAB software controls each aspect of the measurement for rapid and repeatable measurements. The control program operates as depicted in Fig. 4.1. Each instrument in initialized and the user in queried to provide the control parameters for the measurement including dimensions, step sizes, operating frequency, and integration time. The user is then given the option to calibrate in which they would manually connect three calibration standards: short, open, and load. Next the DUT is scanned to the start position and the user is given the option to planarize. If planarization is selected the DUT is automatically scanned to three corners of the scan area and an automated touchdown procedure begins. The corner heights are recorded and a planar height compensation variable is defined. Alternately

optical profilometry is enabled and the DUT height is measured at each location in the scan and recorded for later in-scan compensation. Finally the DUT is moved back to the start position and the scan commences. The mechanical motion stages (x, y) step as specified in the scan parameters and if planarization was conducted the software calculates the necessary height compensation and the piezoelectric motion stage is offset accordingly. The measurement is then taken (either a frequency sweep or a single frequency point), and the result is displayed on the screen. After the DUT has been scanned over the entire area all the data is saved and any desired signal processing is performed.

4.1.1 Differentiation

A differentiation measurement is used to differentiate large quantities of samples (e.g., screening/binning) or classify different materials. This is often a binary detection and should be accomplished rapidly [11]. Thus, a single measurement per sample is desirable, but often better classification can be achieved with additional datapoints from a vertical probe scan over a single (x, y) location. Lateral scans are used to differentiate regions on the same sample.

Figure 4.2 shows a touchdown measurement (vertical scan) over a metallic plane, with and without silicon nitride (SiN) passivation. Circular markers indicate the touchdown over bare metal, and square markers indicate the touchdown over metal with $0.8\,\mu$m of passivation ($\epsilon_r = 6.8$). The probe tip diameter is $5\,\mu$m so we expect the sensitive region to also be around $5\,\mu$m. Indeed, beyond a scan height of $4\,\mu$m the two samples are indistinguishable, but as the probe approaches the passivation layer the difference becomes more and more pronounced. As the tip approaches the passivation the effective permittivity increases resulting in a rapid increase in capacitance which enables effective differentiation. Differentiation of several materials through vertical scanning has also been demonstrated in [62, 63].

Another test structure is scanned for lateral differentiation. Figure 4.3a shows the height profile of a $90\,\mu$m square metallic test structure partially surrounded by SiN, which is destructively measured with a DekTak profilometer. Metallic and SiN regions are indicated, based upon the

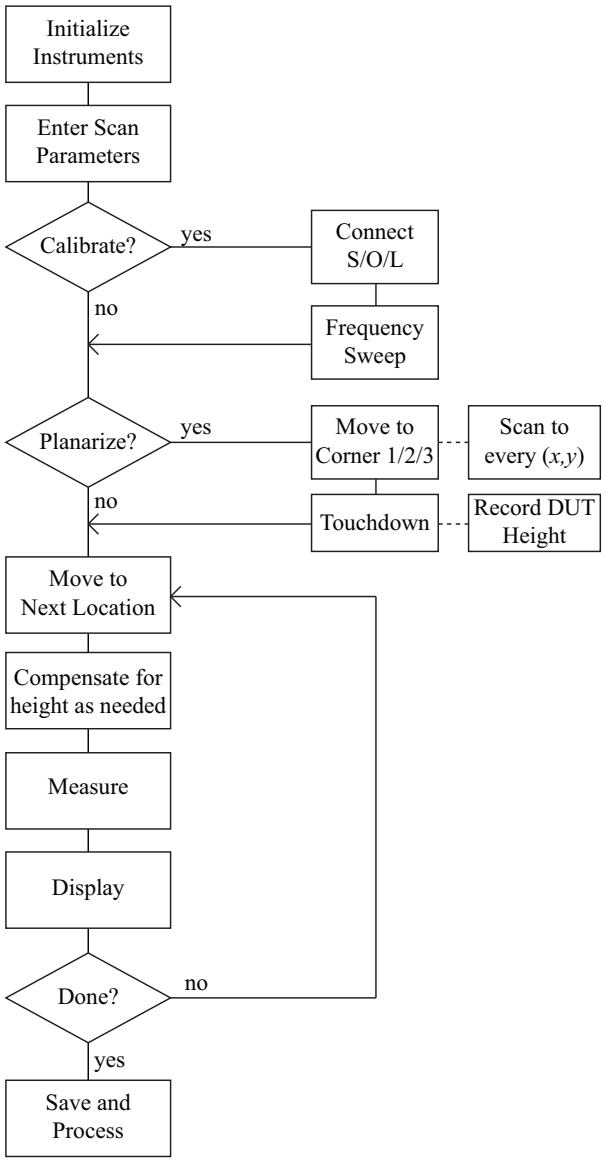


Figure 4.1: MATLAB control software flow graph.

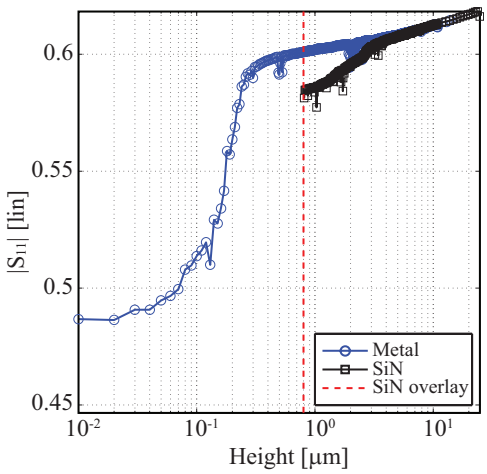


Figure 4.2: A touchdown measurement over metal and over SiN passivated metal yields a measurement contrast.

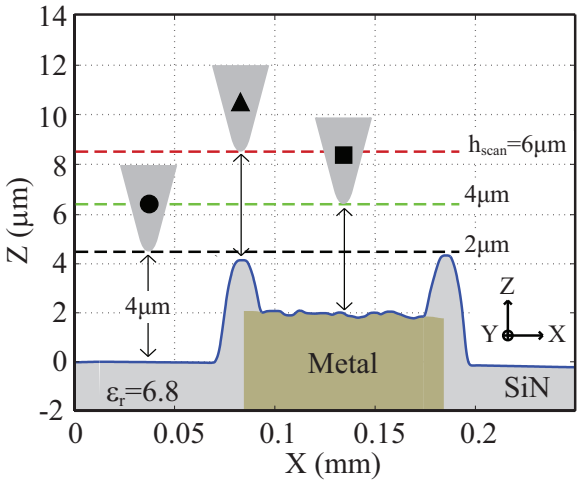
test structure layout, showing that SiN passivation wraps over the sides of the metallic pad. Figure 4.3b shows horizontal scans in X and Y with a 5 μm probe tip at a height of 4 μm above the bare metal. The scans include the expected rectangular dip as the tip transitions over the metal pad, and a response due to buildup of SiN at the edge transition (circled).

To explore the buildup of SiN, several scans are conducted around the edge transition at increasingly close scan heights as shown in Fig. 4.3c. While a step transition from SiN to metal results in a monotonic transition, the non-monotonic response here indicates not just differentiation of SiN and metal but a combination of the two, a much more subtle difference. To ensure we measure only material differences and not topography, we mark symbols in Fig. 4.3c corresponding to a consistent 4 μm height above the sample for each of the three regions along the scan. The height-compensated probe over SiN, SiN/metal, and bare metal results in measured responses of 1820 μV, 1818.25 μV, and 1817 μV respectively, and the RMS noise voltage is approximately 100 $\mathrm{nV_{rms}}$ or 536 $\mathrm{nV_{pp}}$ as shown in Tab. 4.1.

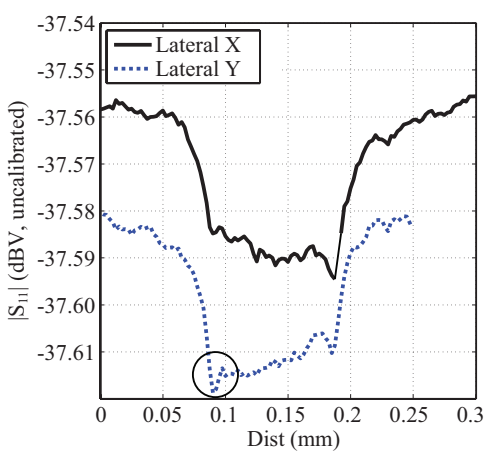
Random variation along the bare metal region in Fig. 4.3b is attributed to surface roughness, also revealed in the height profile, while slope in the sample platform causes the linear bias. If desired, the LVNP can compensate for platform slope by either method discussed in section 4.1. A wider scan of Fig. 4.3c is shown in Fig. 4.4 where the slope is more obvious.

Table 4.1: Signal and noise in differentiation

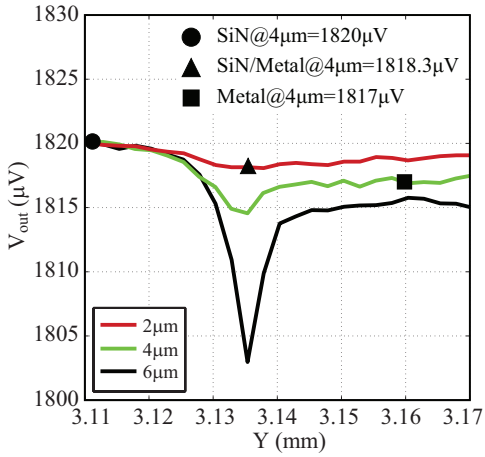
Sample	$\mu[\mu V]$	$\sigma[\mu V]$
SiN	1820	0.1
SiN/Metal	1818.3	0.1
Metal	1817	0.1



(a) DekTak profile of SiN over square metal bond-pad.



(b) $|S_{11}|$ full lateral scan.



(c) $|S_{11}|$ of SiN transition.

Figure 4.3: (a) Height profile of a 90 μm square metallic bond-pad on a MMIC (expanded ordinate scale), destructively measured with a DekTak profilometer. (b) $|S_{11}|$ from a lateral scan with a 5 μm diameter probe tip 4 μm over the bare metal bond pad. Strong responses at the SiN/metal transition (circled) indicates sensitivity to SiN ($\epsilon_r = 6.8$), metal, and a combination of the two. (c) Lateral scans over the SiN/metal transition at several heights ensure material differentiation, not due to topography. Markers indicate the response at a constant height of 4 μm over all three regions.

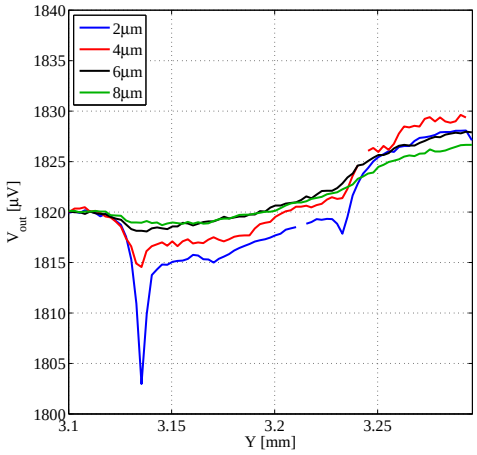


Figure 4.4: Lateral scans at various heights over a metallic pad with SiN overlay built-up on the edges.

4.1.2 Analysis

The analysis modality if used for two purposes in this section: (1) to measure the sensitivity of a given probe as a function of height. The height sweep simulates a continuous range of tip capacitances and thus enables a measure of the probes sensitivity to all possible capacitance values. (2) To model calibrated tip-DUT impedances.

If one is interested in (1) quantifying the sensitivity of a particular probe circuit design then as mentioned, a liftoff measurement is an ideal test because it simulates a large range of possible tip capacitances. This measurement is useful for determining the range of scan heights in which the probe is most sensitive. That is, at what height (or equivalently capacitance) does the tip-DUT impedance change most rapidly. This is known approximately based solely on the diameter of the tip (see chapter 2) but parasitics and the surrounding environment, including the particulars of the DUT will also contribute to the precise result. Figure 4.5 shows a liftoff measurement using piezo modulation. The three traces correspond to modulating the piezo height by 100 nm, 1 μm and 5 μm peak-to-peak. The probe tip diameter used in this measurement was 20 μm and the sample was a metal plane. Because the output of the LVNP is the difference between the two modulation states we expect the amplitude of the 5 μm modulation measurement to be largest. The three traces have the same general shape but close inspection reveals different slopes in

different regions indicating that the piezo modulated measurement is not linear with modulation depth. In addition, and as mentioned in section 3.3.5, the piezo modulated measurement cannot be calibrated. For both of these reasons it is therefore not useful for quantitative modeling and only useful for relative measurements. However, the purpose of this measurement is not modeling but qualification of a 20 μm probe tip. As shown it is indeed sensitive below 20 μm but the SNR begins to degrade and the signal saturates beyond 20 μm. It is not clear from this measurement why the 1 μm measurement is good down to 40 μm.

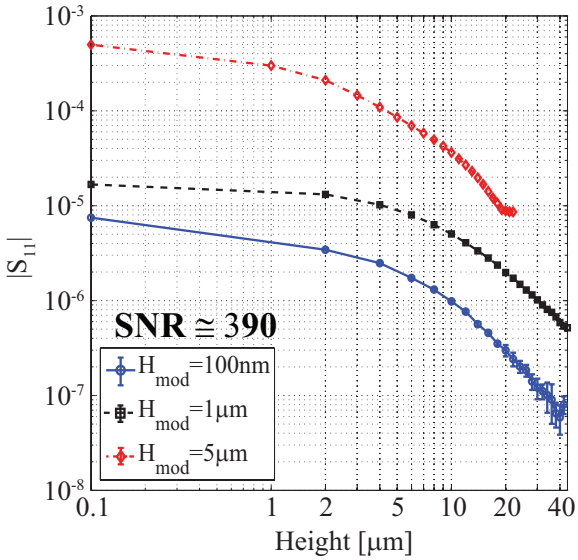


Figure 4.5: Piezo modulation results in higher SNR, but is uncalibrated, thus no meaningful tip capacitance can be extracted.

With LO modulation the probe can be calibrated and (2) qualitative modeling is possible. Figure 4.6 shows the same liftoff measurement but with LO modulation and calibration, and now the data can be modeled and fitted. The figure shows the measured (symbols) and fitted phase response at three heights (10 μm, 20 μm, and 30 μm) in a liftoff scan across frequency. The full frequency sweep is used to assist in more robust fitting than just a single frequency point. The phase is centered around zero at resonance as expected in a calibrated measurement. The extracted calibrated capacitance values are shown in the inset. Though the values are approximately equal to those expected from Fig. 2.15, the trend is incorrect and the height sweep is very coarse.

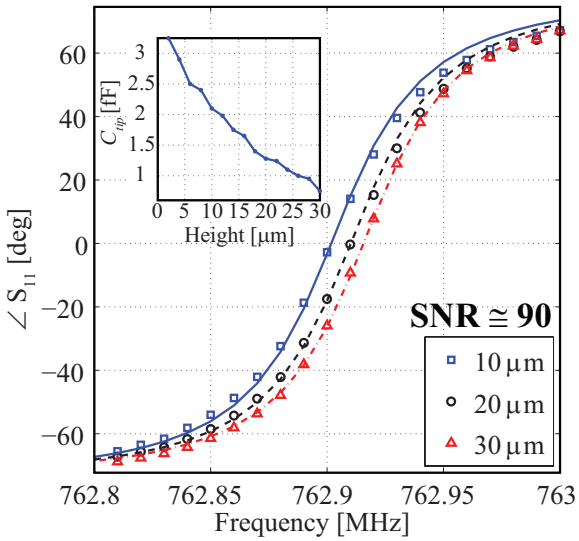


Figure 4.6: The tip-to-sample coupling capacitance can be extracted from LO modulation because of calibration.

As discussed in section 2.6.2 a large stray capacitance is expected due to parasitics including the free space capacitance of the tip C_∞ and the tip shaft capacitance. The analytical models developed in section 2.6.1 do not account for stray capacitance due to parasitics (only accounting for C_∞) therefore we will simulate tip-DUT impedance with FEM simulations as proposed in section 2.6.2. After calibrating a frequency sweep with the probe tip in free space (i.e., far from the DUT such that coupling is negligible) then fitted probe circuit parameters are de-embedded to reveal the complex coupling impedance of the tip-DUT at the P2 reference plane. Finally, finite element method (FEM) electromagnetic simulations (Fig. 4.7) are used to account for the extra tip parasitics as in [64].

Figure 4.8 shows a very precise touchdown measurement over metal with a 5 μm diameter probe tip, 10 nm step sizes, and $\tau_{\text{int}} = 3$ seconds for good SNR. Applying SOL calibration, then de-embedding the probe circuit model, the ordinate axis is converted from an uncalibrated $|S_{11}|$ to a calibrated tip capacitance, C_t. Also shown in Fig. 4.8 is the theoretical capacitance of a parallel plate capacitor with circular plates of radius 10.8 μm (or twice that of the spherical tip as expected from section 2.6.1). However, outside a small region around 200 nm the capacitance diverges from this ideal case because it does not account for stray capacitance or the complicated

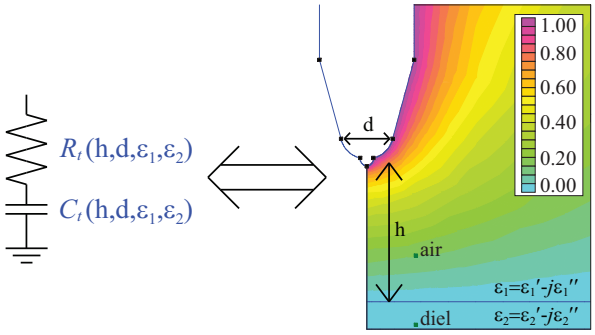


Figure 4.7: 2D FEM electromagnetic simulation of the electric potential distribution around a probe tip of diameter d, a height h off a slab of dielectric over metal is used to develop equivalent circuit models of tip-DUT loading from calibrated and de-embedded measurements as in Fig. 4.8.

tip geometry. The dashed trace is the capacitance of the shaft and taper of the probe tip as calculated from FEM simulations as in Fig. 4.7, showing that the saturated response further from the sample is dominated by the shaft. If the shaft length is halved the saturated response (dash-dot trace) lowers by approximately 2.5 fF which would extend the sensitive region of operation by 100 nm. The response close to the sample (< 100 nm) approaches a non-singular limiting value as expected from (2.36), with deviations from theory probably due to inevitable small imperfections at the apex of the spherical tip [45].

4.1.3 Imaging

The final modality is imaging which results in a 2D image over the lateral coordinates of the DUT using a constant step size throughout the scan. The 2D image contrast is a map of the variation in composition of the DUT. One goal of the LVNP is to provide very large scan range compared to the often implemented AFM-assisted NFMM systems [65, 7]. The scan area in most of the following images is typically 1 mm×1 mm, which is 100 times larger than even an extended range AFM scan head [42].

Figure 4.9a shows an optical image of a modern CMOS integrated circuit (IC) fabricated in the IBM MOSIS process. The black square shows the scan region of interest with two 10 μm-wide bus lines shown conceptually. The zoom-image shows detail of two layers of metallic fill at the surface of the IC, covering the parallel bus lines. CMOS fabrication often has a metal fill requirement

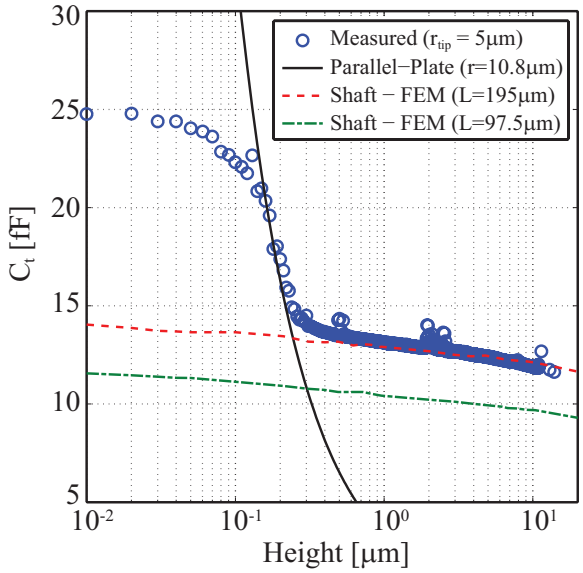
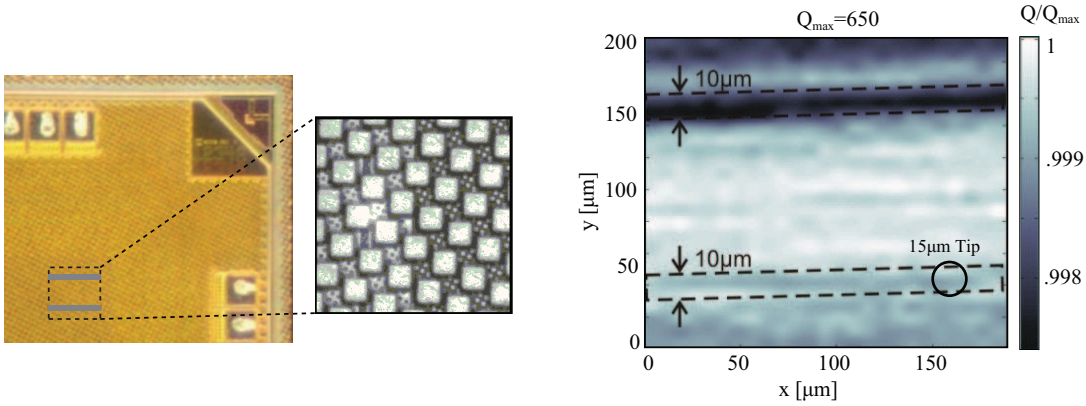


Figure 4.8: Three distinct regions exist in a typical touchdown measurement. Below approximately $0.1\,\mu$m small imperfections at the tip of the probe tip dominate the response. Beyond $0.5\,\mu$m the taper and shaft of the probe tip dominate the response effectively saturating the sensitivity. Between these two regions the tip can be effectively modeled as a parallel-plate capacitor. When the shaft length is halved the saturation value decreases.



(a) Photo of CMOS IC.

(b) $Q/Q_{\max}$.

Figure 4.9: (a) Photograph of a CMOS IC with metal fill (zoom). Scan area is indicated in the black square. (b) Normalized Q-factor of the probe circuit with a 15 μm probe tip showing detection of a sub-surface 10 μm-wide bus line. $Q_0 = 650$ and the measured change here is 0.3%. The near-field tip can interact with the bus lines only because the metal fill is spaced with dielectric as shown in the zoom of (a).

that each layer must consist of at least 30 % metallization in order to assist in layer-to-layer planarization. The goal of this measurement was to detect[1] two 10 μm-wide parallel bus lines below the two layers of metal fill. It is not expected that the fields from the tip would interact with a structure below two layers of metal fill except that, as shown in the zoom of Fig. 4.9a, the fill is a grid of metal squares with dielectric filling the interstices, between which the bus lines do load the near-field probe tip. Fig. 4.9b shows the normalized Q-factor ($Q/Q_{\max}$, where $Q_{\max} = 650$) of the probe circuit over the scan area with a 15 μm probe tip. Though this was an early measurement with significant drift and excess noise in the LVNP there is still a strong response in the region of the bus lines as shown in the dashed outline. The rest of the image is an average response of the fill with noise and drift. This measurement shows not only 2D imaging but sub-surface detection. As will be evident in the following 2D images, after the noise and drift issues were addressed (as discussed in section 3.3.2) the measurements achieved much higher SNR and less drift across the scan area.

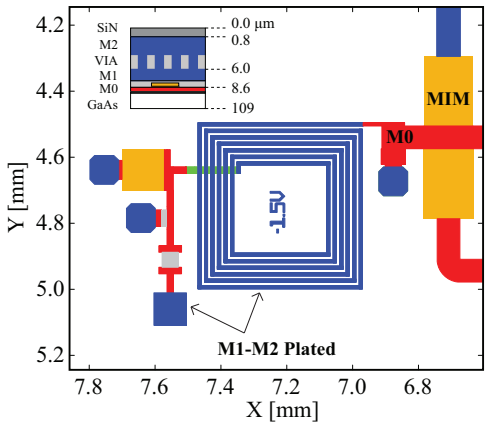
Figure 4.10a shows the layout of a second test structure fabricated in the TriQuint TQPED GaAs MMIC process, with three layers of metallization. M0 is the lowest layer, 8.6 μm below

[1] Detection differs from imaging in that features may not be spatially resolved according to the Nyquist criteria but the sub-resolution features have an effect on the signal level within a given pixel.

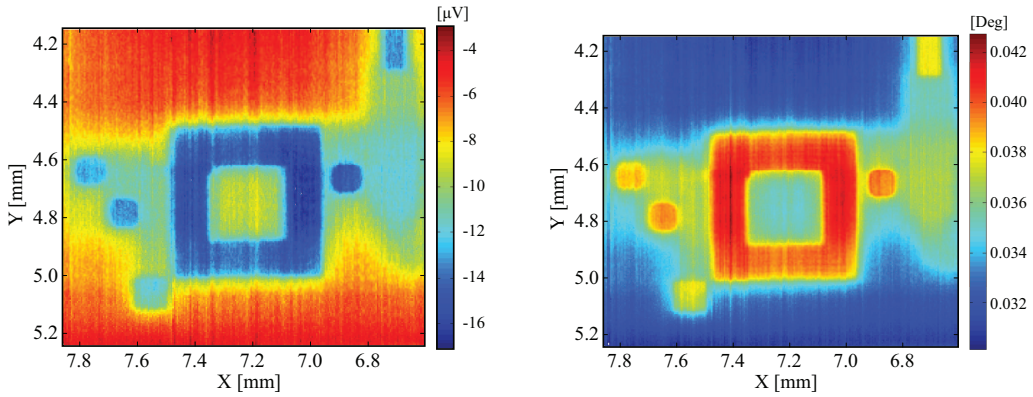
the surface passivation. A MIM capacitor layer is 0.15 μm above M0. M1 is 6 μm below the surface but is via-plated to M2 which lies 0.8 μm below the surface passivation (referred to as M1-M2). The planar spiral inductor in the middle of the MMIC has 10 μm lines and spaces and is composed of plated M1-M2. Figures 4.10b and 4.10c show the magnitude and phase response from a 20 μm probe tip scanned over the test structure. The various metallic layers are clearly differentiated with the strongest response coming from the thick plated metal layer (M1-M2) near the surface and the lower resolution diffuse response coming from the lower metal layer (M0) buried beneath nearly 9 μm of dielectric. $|S_{11}|$ has a minimum value when it is most strongly loaded by the plated metal because as the capacitance increases, the resonance shifts down. $\angle S_{11}$ has the opposite trend because the phase of the response increases around resonance (see positive slope of phase around zero in Fig. 4.6). Despite drift-induced vertical stripes which are more pronounced in $|S_{11}|$, it is useful to examine both magnitude and angle because of the opposing trends. We point out the extremely good stability across the entire scan area even though this measurement required approximately 2 hours to complete.

Figure 4.11 shows the same test structure scanned with a large probe tip (50 μm) and a small probe tip (10 μm). As expected, features are significantly blurred with the large probe tip but an overlay of the scan with the small tip shows the 10 μm lines and spaces of the inductor and the octagonal shape of the probe pad. This confirms that spatial resolution is on the order of the probe tip diameter.

Instead of characterizing spatial resolution with lines and spaces of decreasing dimension, we can alternatively analyze the response of an edge target. Figure 4.12a shows the edge response or edge spread function (ESF) of a 10 μm tip scanned over plated M1-M2, a 10 μm tip scanned over the sub-surface M0 and a 50 μm tip scanned over plated M1-M2. Scan tracks over the test structure are indicated in the inset. From the ESFs we can qualitatively say that the 10 μm tip over M1-M2 has the steepest rise and thus the highest spatial resolution, while the same tip over the buried M0 edge shows more noise and a shallower rise. The 50 μm tip over the M1-M2 edge shows the expected shallower rise as well. These observations can be further quantified through



(a) Target.



(b) $|S_{11}|$. (c) $\angle S_{11}$.

Figure 4.10: (b) and (c) show the magnitude and phase of the reflection coefficient of the probe circuit as it is scanned over a 1.1 mm by 1.0 mm area of a MMIC inductor circuit as shown in (a).

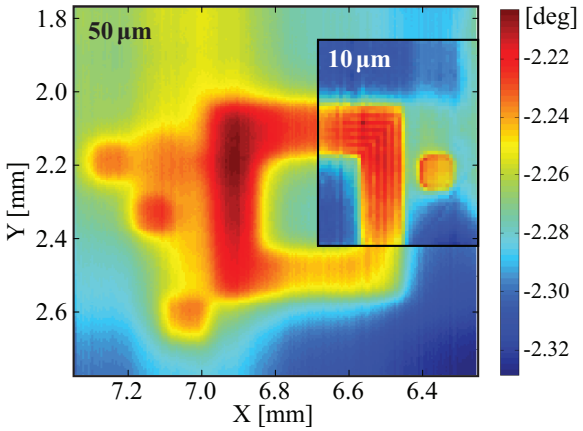


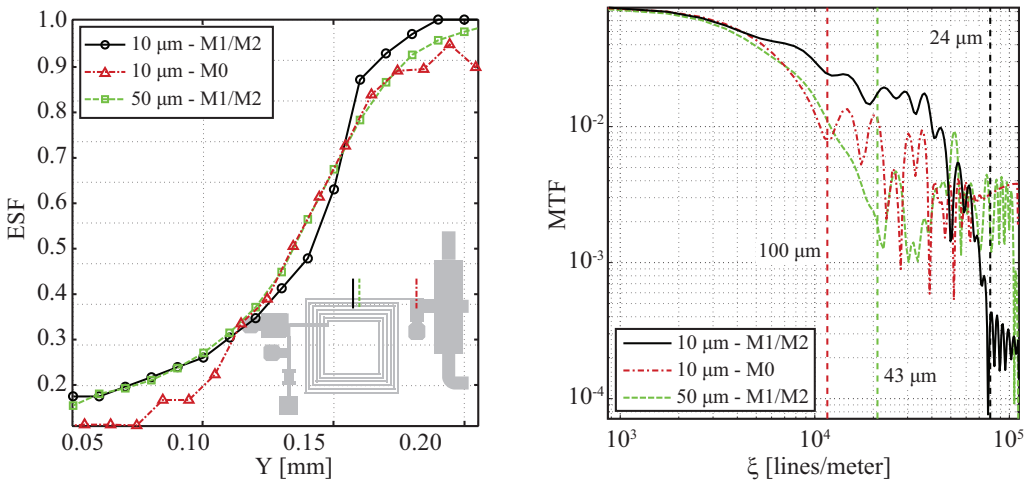
Figure 4.11: The MMIC test structure is scanned with a large 50 μm probe tip and a small 10 μm probe tip (overlay). The large tip blurs features revealed by the smaller tip including the 10 μm lines and spaces of the inductor and the octagonal shape of the bond pad.

fourier analysis and the optical transfer function (OTF).

The OTF describes the magnitude and phase of the system response as a function of spatial frequency ξ. It is the Fourier transform of the derivative of the ESF [66], and the modulation transfer function (MTF) is the magnitude of the OTF. The edge responses shown in Fig. 4.12a yield the MTFs shown in Fig. 4.12b. Spatial resolution is defined as the point at which the MTF drops to the noise floor and is indicated by vertical dashed lines for each of the three measurements. From the MTFs we can quantitatively say that the 10 μm tip over M1-M2 exhibits 80000 line-pairs/meter spatial resolution, or 24 μm resolution. This degrades by a factor of four to 100 μm over M0, as a result of the 9 μm dielectric covering. The 50 μm tip over M1-M2 has 43 μm resolution. The low resolution of the 10 μm tip could be due to a blunted tip, while the high resolution of the 50 μm tip is likely due to small scale tip imperfections [45], and better SNR due to the large tip signal.

4.2 Summary

In this chapter, several measurement modalities were defined. In each modality measurements were presented and analyzed. Calibration and signal processing algorithms were described as appropriate. Differentiation of dielectric, metal, and a stackup of both was achieved. Calibrated tip coupling capacitance was measured which agreed well with analytical and FEM predictions. Imaging scans demonstrated extremely good stability through a 2 hour scan, and spatial resolution was demonstrated and quantitatively analyzed as a function of tip-diameter, and sample depth. The contributions in this chapter are presented in [58, 20, 60].



(a) ESFs

(b) Corresponding MTFs.

Figure 4.12: (a) Edge spread functions of the transition from dielectric to metal (scan tracks shown in inset) with a 10 μm and a 50 μm tip over plated M1-M2 and over sub-surface M0 are processed to yield (b) MTFs which show the quantitative response as a function of spatial frequency. The MTFs confirm the smaller probe tip provides higher spatial resolution but when imaging metal below dielectric, at a depth on the order of the tip diameter (here 9 μm) the spatial resolution degrades by a factor of 4.

Chapter 5

Conclusions and Future Work

Contents

5.1 Summary & Contributions

In this work we have presented the design, analysis, and characterization of a near-field probe tip, microwave probe circuit, and the LVNP instrument for the non-destructive microwave

measurement of micro- and nano-scale samples including inhomogeneous materials, MMICs, and complex structures with lateral and vertical composition. We discussed the concept of sub-wavelength spatial resolution, as well as various tip-DUT models. Then we presented the design of a high Q-factor, critically-coupled resonant probe circuit, and a low-noise measurement system (the LVNP) for modulation and phase sensitive detection of NFMM measurements. The system noise floor was characterized and noise limiting elements were identified. The $1/f$ knee was measured as $f_c \leq 0.01\,\text{Hz}$, at least four orders of magnitude better than the same electronics without modulation. This stability enabled large scan-area measurements with acceptable levels of drift. It was shown that the LVNP can differentiate not only dielectrics and conductors, but similar dielectrics. Through custom SOL calibration standards and FEM simulations the physical tip impedance was determined. Further it was found that sensitivity could be increased by shortening the shaft of the tip. Images of a CMOS test structure showed sub-surface detection, and images of a MMIC test structure showed good long-term stability as well as spatial resolution on the order of the probe-tip diameter. A quantitative definition of spatial resolution was based upon the modulation transfer function as calculated from edge targets and spatial resolution as a function of sample depth was discussed.

Here we summarize the various contributions of this thesis from each chapter, with references as appropriate:

5.1.1 NFMM Concept and Probe Circuit (Ch. 2)

- Sub-wavelength spatial resolution was justified through the angular spectrum expansion.
- High-Q, critically coupled probe circuit designs were motivated and implemented.
- An accurate and simple probe circuit model was developed for design and simulation, and later for data extraction.
- Several tip-DUT modeling methods were presented.
- An efficient method of extracting tip-DUT coupling through MATLAB automation of

FEMM simulations was presented.

These contributions are presented in [21].

5.1.2 Lock-in Vector Near-field Probe (Ch. 3), & Low-noise THz Imaging Readout (App. A)

- Generalized modulation for the reduction of white and $1/f$ noise in terahertz (THz) imaging was discussed within a mathematical framework.
- Generalized modulation concepts, or PSD were applied to the LVNP and include both calibrated LO modulation and low-drift piezo modulation.
- A simple and effective calibration method was applied to the LVNP for the extraction of calibrated tip-DUT impedances. This, together with tip models and FEMM fitting provided quantitative tip capacitances.

These contributions are presented in [58, 59, 60].

5.1.3 Measurements and Analysis (Ch. 4)

- Defined three measurement modalities including differentiation, analysis, and imaging.
- Differentiation of dielectrics, and quantitative modeling of tip capacitances for the optimization of tip geometries.
- Measured calibrated tip capacitance which agreed well with analytical and FEM predictions.
- Presented an appropriate image processing method for accurate analysis of spatial resolution through the modulation transfer function (MTF).
- A quantitative examination of spatial resolution as a function of depth, and material.

These contributions are presented in [60].

5.2 Future Work

5.2.1 Integration and Parallelization

The LVNP presented in this work can be extended in several ways. The probe circuit was designed in a planar topology [21] for repeatable fabrication and simple integration of various probe tips. In the future the planar probe circuits and the LVNP instrument could be further miniaturized and integrated into arrays of multi-channel probes. Arraying has two benefits: (1) arrays of N identical elements provide N-times faster scans, and (2) multi-frequency arrays provide spatial resolution at various depths [23]. Toward this end several MMICs have been or are under design. The LVNP *IQ*-demodulator has been implemented as a pair of quadrature singly-balanced gallium-arsenide (GaAs) MMIC mixers providing good LO-IF isolation with low LO power requirements. A quasi-circulator [67] and LNA design are under way to replace the ferrite circulator. Even the surface-mount high-Q resonator can be replaced by MMIC-integrated resonators at the expense of Q-factor (from $\sim$1500 to $\sim$400) [68, 69]. Gains achieved by locating the readout at the probe tip could partially compensate for the lowered signal due to decreased Q-factor. Multi-channel arrays will require multi-channel lock-in amplifiers, which we have previously demonstrated with up to 128-channels in a chassis of printed circuit boards (PCBs) [59]. In the following we briefly overview progress regarding each of the extensions mentioned.

MMIC IQ-demodulator

The *IQ*-demodulator in the LVNP serves as a homodyne receiver for acquiring magnitude and phase of the reflected voltage from the probe circuit. In Fig. 3.10 it was shown to be the noise-limiting component due to the LO amplifiers in the commercial chip we are using. Mixers often include LO amplifiers to ensure the non-linear mixer core(s) operate in the strongly non-linear region. As an alternative to amplifying the LO it is common to bias the mixer cores. However, both of these methods for achieving low conversion loss are not suitable for the LVNP as the LO amplifier will contribute to the noise floor and cause drift, and biasing the mixer cores also causes drift. For these reasons we designed a fully passive *IQ*-demodulator in the TriQuint TQPED

GaAs MMIC process with a design frequency of 1500 MHz. While the TQPED process includes Schottky diodes, an obvious choice for the non-linear mixer core(s), it was found that diode-connected FETs produced a sufficiently strong non-linearity at lower bias power levels and would thus enable lower LO drive levels. A balanced mixer architecture is desirable for its inherent RF and LO input match, and suppression of some of the intermodulation distortion (IMD) products. Because of the narrow-band homodyne application a hybrid power divider is used to implement the balanced architecture, but at the low operating frequency lumped-element equivalent circuits must be used. A single-balanced mixer, while canceling only the even IMD products and with lower RF-LO isolation than a double balanced mixer, generally requires half the LO power and still provides sufficient isolation and matching for this application. Therefore the mixer cores are in a single-balanced configuration as shown in Fig. 5.1 to provide inherent LO-RF isolation, LO and RF input match, and suppression of LO amplitude fluctuations. Lumped-element filters are included throughout the design to provide additional forward and reverse isolation. A schematic of the final layout and photograph of the fabricated circuit are shown in Fig. 5.2.

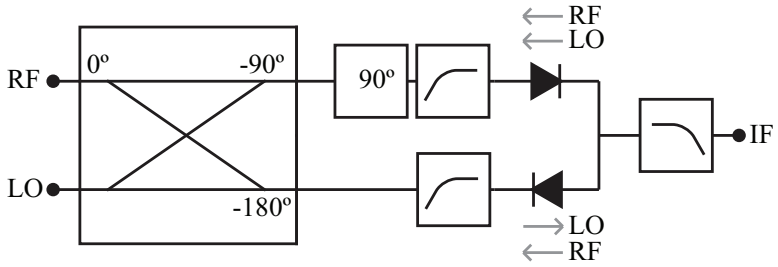
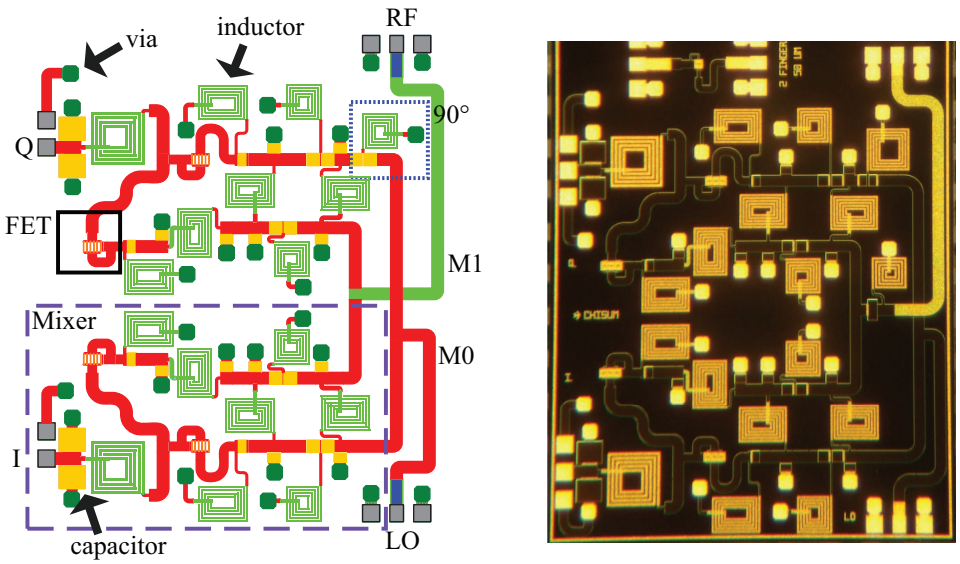


Figure 5.1: A single-balanced mixer consisting of a quadrature hybrid, a 90° phase delay, and inverted mixer cores provides inherent input match, LO-RF isolation, and rejection of even IMD products.

A mixer spreads input power over the frequency domain. Ideally the power splits to just the sum and difference frequencies and the mixer has a 3 dB conversion loss. However, the mixer cores produce intermodulation products and thus the mixer always incurs several dB more conversion loss. The passive mixer cores ideally switch between infinite and zero impedance at the LO frequency, however, this is never the case and so there are dissipative losses during the switching of the cores which will also contribute to conversion loss. Finally, any reflections at the RF and



(a) IQ-demodulator layout.

(b) IQ-demodulator photo.

Figure 5.2: The IQ-demodulator is comprised of two single-balanced mixers in phase-quadrature.

LO input ports, as well as losses in the lumped elements also contribute to conversion loss. In total we expect no less than 8–10 dB conversion loss. In fact, the measured conversion loss in Fig. 5.3 is shown to be typically 11.5 dB at $P_{\mathrm{LO}} = 14\,\mathrm{dBm}$ and 13 dB at $P_{\mathrm{LO}} = 10\,\mathrm{dBm}$. The increased conversion loss is attributed to additional losses in the lumped-element equivalent circuits which were not fully accounted for in the linear simulator. Full-wave simulation of the lumped-element hybrid revealed an additional 1 dB of loss than predicted by the linear simulator. Finally, non-ideal RF/LO chokes shunt power to ground thereby increasing conversion loss.

By nature of the balanced architecture and matched mixer cores the RF and LO ports have a measured input match of better than −15 dB at the design frequency of 1500 MHz with RF power at −20 dBm over a range of LO power from 0 dBm to 10 dBm as shown in Fig. 5.4. Because the IQ-demodulator is to be used in a homodyne configuration where the LO power level and phase will be fixed, the output is then a function of only the RF magnitude and phase:

$$V_I \propto V_{\mathrm{RF}} \cos\left(\Theta_{\mathrm{RF}}\right) \tag{5.1}$$

$$V_I \propto V_{\mathrm{RF}} \sin\left(\Theta_{\mathrm{RF}}\right). \tag{5.2}$$

Figure 5.5 shows the output as a function of phase for fixed LO and RF power levels. For

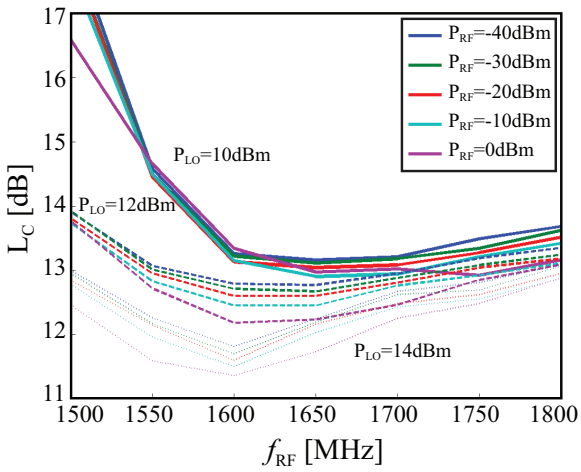
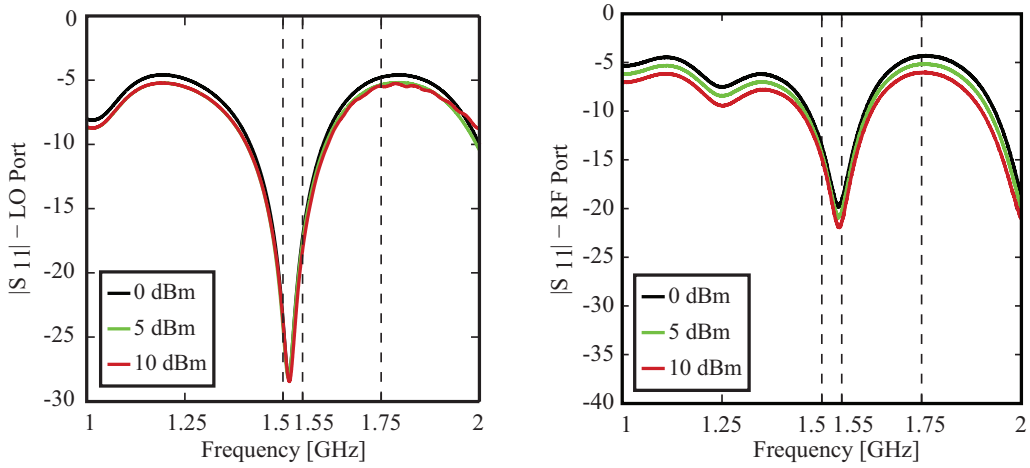


Figure 5.3: For 14 dBm LO power, conversion loss is approximately 11.5 dBm at 1600 MHz. For lower LO powers conversion loss increases by 1 dB per 2 dB decrease in LO power.



(a) LO match.

(b) RF match.

Figure 5.4: (a) The RF port is well matched at 1.55 GHz, 50 MHz off from the design frequency. (b) The LO port is well matched at the design frequency of 1.5 MHz.

$P_{RF} = -20$ dBm and $P_{RF} = -10$ dBm (the expected range) the output is the cosinusoid we expect from (5.1). However, as the RF power increases to 0 dBm, the mixer cores are driven harder and the output becomes non-linear which distorts the waveform.

MMIC Quasi-circulator

The forward and reflected waves to/from the NFMM probe circuit are separated by a ferrite circulator as shown in Fig. 3.2. The circulator is inherently narrow-band, bulky, and requires

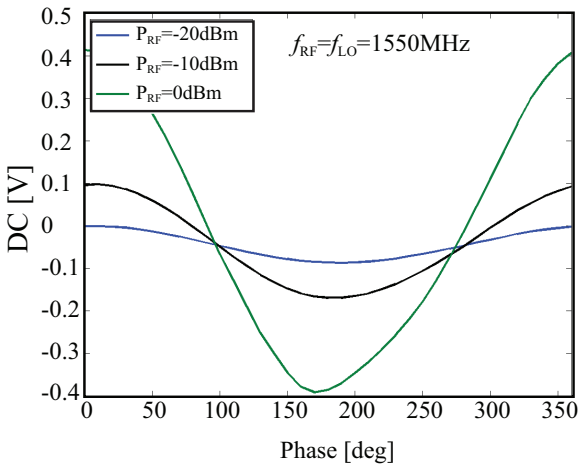
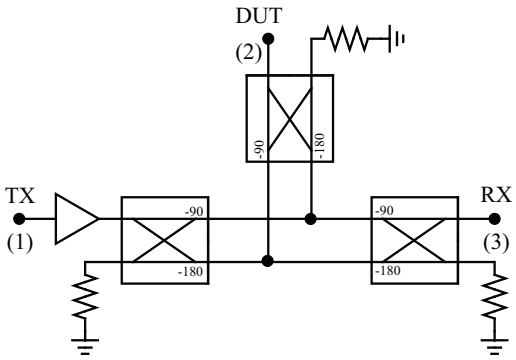


Figure 5.5: The expected sinusoidal output of the mixer configured as a phase detector degrades for RF power of 0 dBm due to non-linearities in the mixer cores. At and below RF powers of −10 dBm the phase detector works as expected with a slight offset of the origin.

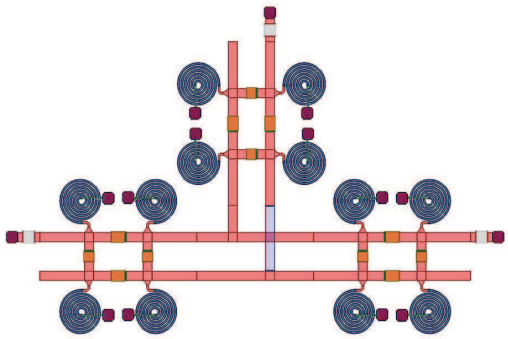
a ferrite which is not compatible with standard MMIC processes like the TriQuint TQPED. Alternative methods are available for separating forward and reverse waves, however the standard directional coupler does not provide sufficient isolation. A quasi-circulator can provide high isolation and MMIC integration by relying upon phase cancelation and non-reciprocal amplifiers. Quasi-circulators differ from circulators in that $S_{13} = 0$ and $S_{23} \neq 0$. That is, port 3 does not circulate to port 1, but does reflect back to port 2. However, the IQ-demodulator will be connected to port 3 and has already been shown to have a good input match so the failure of the quasi-circulator is not an issue in this application. Shown conceptually in Fig. 5.6a, it relies upon three hybrids configured in such a way that phase cancelation provides isolation from port 1 to port 3 ($S_{31} \approx 0$). However, power from port 2 is equally split to port 1 and port 3 so we require an asymmetric device at port 1 to provide isolation ($S_{12} \approx 0$). A layout of the triple hybrid is shown in Fig. 5.6b and the amplifier is yet to be designed.

Probe Arrays

Once the LVNP is implemented in MMICs the system can be integrated onto the probe circuit itself. In fact, multiple LVNP MMICs can be implemented on one small circuit board paving the way for arrays of probes. Figure 5.7 shows two concepts of probe arrays: an array of N identical



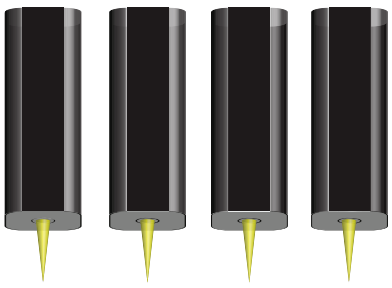
(a) Quasi-circulator concept.



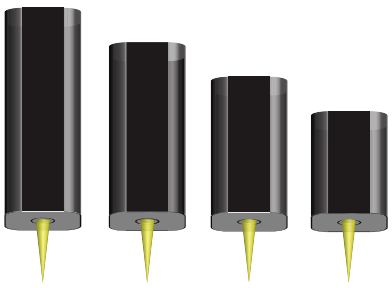
(b) Phase-canceling hybrids.

Figure 5.6: (a) Through phase cancelation and an asymmetric amplifier, a quasi-circulator can be constructed without ferrites. (b) Three lumped-element quadrature hybrids are configured for phase cancelation.

elements (a) for an N-fold decrease in scan time, and an array of N multi-frequency probes (b) for multispectral measurements. According to the skin-depth (see section 2.6.2) multi-frequency probes will image at different depths within a sample which paves the way for advanced image processing like reverse-tomography. Note that probe arrays fundamentally differ from antenna arrays in that the entire array is concentrated into a volume much less then a wavelength in radius. As such, array factors are meaningless, beam steering is not possible, and coupling must be considered. For this reason we suggest the use of guided wave structures and resonators that result in strongly confined fields.



(a) N-element array.



(b) Multi-frequency array.

Figure 5.7: (a) An array of N identical probes will increase scan speed by a factor of N. (b) An array of N probes of different frequencies provides a multispectral dataset. With a parallel readout, the scan time for a multi-frequency probe array is the same as a single element scan except that each (x, y) location is measured at each frequency.

In order to truly benefit from the advantages of many-element arrays the resonator must be

planarized, and ideally integrated into a MMIC. The resonator for a 762 MHz probe as discussed in chapter 2 is 20 mm long, with a larger footprint than ten LVNPs implemented in a MMIC. It is possible to fully integrate the resonator at the expense of Q as shown in [68, 69] where we would expect the Q to degrade from ~ 1400 to around ~ 400. This reduction in Q will degrade the probe sensitivity but it is possible that integration of the LVNP onto the probe circuit could mitigate some of this by removing cables and connectors and providing matched thermal and interference environments.

Parallel readout

Arrays of probes are only useful if they can be read out in parallel. That is, mechanical scan time is not the speed bottle-neck, it is the required integration time at each location. Therefore, an N-channel array requires an N-channel lock-in amplifier to integrate all probes in the array simultaneously. The multi-channel readout circuitry developed in appendix A is ideally suited for this purpose and a photograph of the final circuit is repeated here in Fig. 5.8 for convenience.

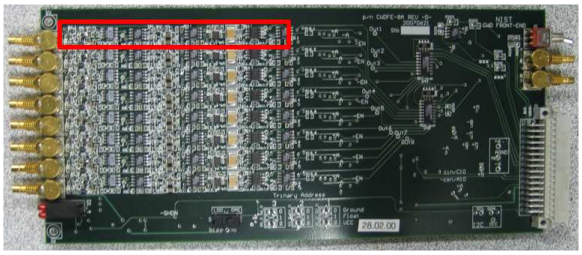
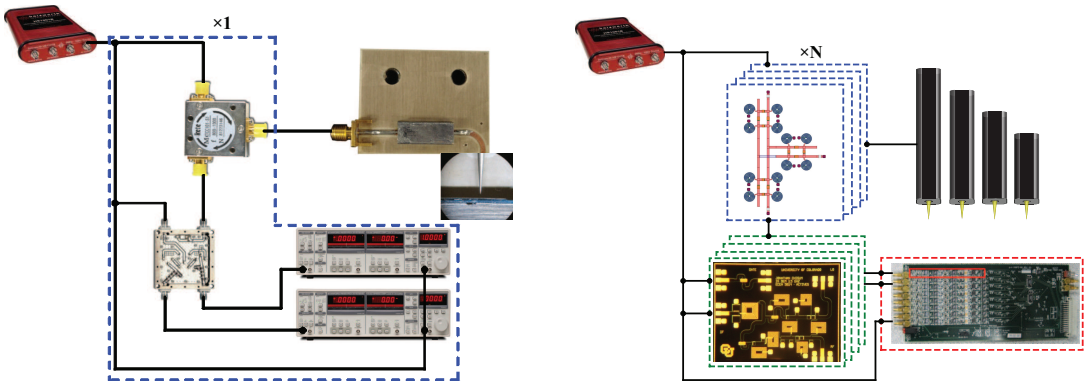


Figure 5.8: Eight lock-in amplifier channels are included on one PCB with up to 16 PCBs in a single chassis allowing for probe arrays of up to 128 elements.

The end result of all this integration is a conversion from the present componentized LVNP and probe circuit, to a fully integrated version with MMICs and arrays of probes. The full conversion is depicted in Fig. 5.9.

5.2.2 Measurements

With the present and evolving LVNP many measurements not discussed in this work are possible. Two are highlighted here: (1) near-field microwave detection (NFMD), possibly combined with



(a) Componentized system. (b) Integrated system.

Figure 5.9: Once the individual instruments and components of the LVNP are replaced by MMICs, the entire system could be integrated onto once circuit and will enable parallelization and arrays.

rapid optical imaging, and (2) magnetic NFMM. Some work has already been done regarding (1), and only minor modifications are required for (2).

Near-field Microwave Detection

As mentioned in chapter 1 there is significant interest in modern IC diagnostic tools. One of the challenges of modern IC diagnostics is determining whether the fabricated chip is as designed. The DARPA TRUST-IC program was initiated to address concerns regarding malicious tampering of ICs during off-shore fabrication. In this scenario one desires to know if one chip is the same as a known good standard which is a fundamentally different problem than imaging, it is a problem of detection. In this scenario one could scan an IC and, given sufficient SNR compare a measured value at an (x, y) location to the standard value. Even if the scanning tip is not capable of spatially resolving the e.g. 22 nm features, the 22 nm features will contribute to the overall signal and thus it is possible to detect a 22 nm change in an IC. Aside from the challenge of SNR, an IC is very large and scanning the entire area with high spatial resolution would require a prohibitive amount of time. As we demonstrated in [21], optical and infra-red laser imaging can be used to identify areas on an IC for finer resolution and/or higher SNR scanning with the LVNP.

Figure 5.10 shows two nominally identical cores (1 and 2) of an IC, differing only in that a small section of circuitry in the lower left corner of core 2 is missing. These images were performed

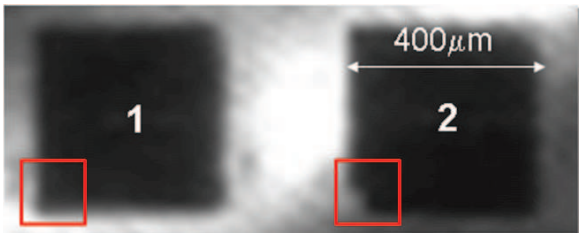


Figure 5.10: Core 1 and core 2 differ in the lower left corner in back-side IR laser imaging measurement.

using back-side infra-red laser imaging. The laser imaging provides a low-resolution video-rate inspection of the entire chip, then regions of interest are scanned in more detail with the NFMM. Even though the NFMM probe tip may be larger than $\lambda/2$ of the laser it has the advantage of arbitrarily high SNR, thus sub-resolution detection is possible. Figure 5.11 shows the identified corner from NFMM. 20 nm features are not resolved but there is an average difference between the two cores. Several scans were taken of each core and the mean and variance was calculated for each location in the scan. Then the mean values from each scan were compared to detect a difference in the lower left corner. The SNR of 2, as shown in Fig. 5.12, was deemed insufficient for reliable differentiation, however the current NFMM has vastly superior long-term stability and a lower noise floor than the earlier version use in these measurements so modern IC differentiation could be more effective in the future.

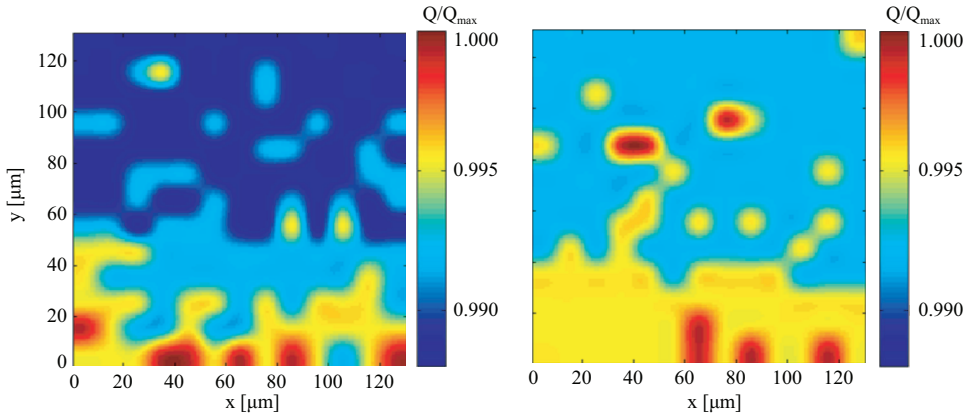


Figure 5.11: near-field microwave detection measurements of the lower left corner of core 1 (left) and core 2 (right). 20 nm features are not resolved but the detected signal at each location measuring a bulk difference.

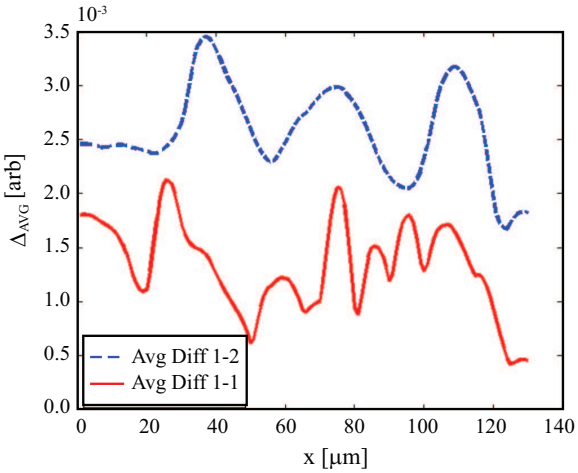


Figure 5.12: Several repeated scans of each core are averaged and then compared against each other. The average difference between subsequent scans or core 1 (red, solid) is half that of the average difference between core 1 and core 2 (blue, dashed). The SNR= 2 is low but suggests differentiation. With the present more sensitive NFMM the SNR would be dramatically improved.

Magnetic field probe

As discussed in chapter 1, the near-field probe can be either an electric probe or a magnetic probe. The probe used in all of this work is an electric field probe in which a sharp conductor serves to localize the electric field. However, an equally viable probe would be a small wire loop loading the probe circuit resonator, as shown at right in Fig. 5.13. The electric field probe circuit was originally designed with a half-wave open-circuit resonator to maximize the voltage at the tip-end of the resonator. In this case the magnetic field should be maximized and thus the resonator should be a quarter-wave with a short-circuit at the tip. In both cases the input impedance is high and so critical coupling should be implemented for the magnetic probe just as it was in the electric probe. The small wire loop probe could be implemented with a length of wire-bond. One significant difference between the electric and magnetic probes is that while the electric probe had rotational symmetry and thus equal x and y spatial resolution, the magnetic probe shown here would have significantly asymmetric lateral spatial resolution.

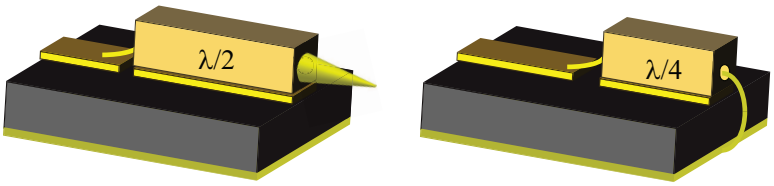


Figure 5.13: An electric field probe should peak the voltage at the tip so the resonator should be a half-wave open, while a magnetic field probe should peak the current at the tip so the resonator should be a quarter-wave short.

Appendix A

Ultra-low-noise THz Imaging Readout

...the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity, and affects not the pomp of superfluous causes.

—Isaac Newton, in *Philosophiæ Naturalis Principia Mathematica, v.III: System of the World*

Contents

A.1 Foreward

We now examine the problem of detector readout in THz imaging arrays. The problem, while distinct from NFMM bears many similarities. Much of the author's prior work in this field was directly applied to significant portions of the design and analysis of the LVNP in chapter 3 including the noise theory. The 8-channel readout circuit described in this chapter is proposed for direct application in section 5.2.1. For these reason we summarize the author's work in low-noise readout of THz imaging arrays here.

A.2 Introduction

Millimeter-wave and terahertz imaging has been shown to be useful for concealed weapons detection through clothing [70], low-visibility guidance [71], and other applications.[72] The availability of sensitive detectors such as bolometers and diodes makes passive imaging practical. The main challenges of passive terahertz imaging are a result of low signal and high noise levels at room temperature. Low SNR can be overcome with long integration times, but this makes video rate imaging[73] difficult. Presently, each imaging system requires a dedicated readout circuit[74] but most terahertz imaging systems suffer from the same problems and thus a general approach to detector readout is useful.

The goal of the first part of this paper is to present a general and comprehensive discussion on requirements of readout circuitry for room-temperature video-rate passive imaging. In order to increase the imaging speed, an array of detectors is used instead of a single, scanned detector. For this reason, the readout electronics consists of multiple channels of phase-sensitive low-noise amplifiers [75, 76], as illustrated in Fig. A.1. Radiation from a blackbody target, characterized by temperature and emissivity, is modulated (at f_{mod}) through a mechanical chopper and received by an antenna-coupled detector array. The modulated nanovolt-level output of each detector is amplified and integrated prior to digitization. The remainder of the paper discusses the design and optimization of the readout electronics which accomplishes phase-sensitive amplification and

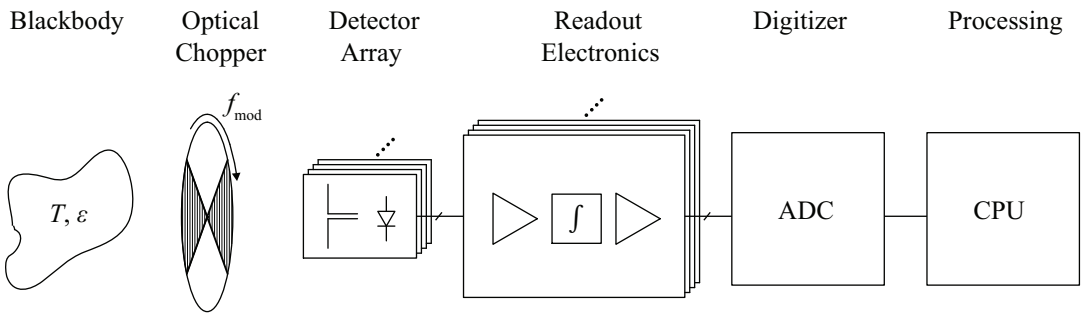


Figure A.1: Radiation from a blackbody target is modulated, detected, measured with a low-noise readout circuit, and digitized.

signal integration of various arrayed detectors.

The two commonly available detectors from around 100 GHz to several THz are antenna-coupled resistive bolometers [77] and semiconductor diodes.[78] A bolometer changes resistance as a function of temperature, which in turn changes due to varying incident RF power. In this work, bolometers are modeled as a voltage source with a low series resistance (hundreds of ohms[79]). Varying incident RF power on a diode is modeled as an induced current source with a high parallel resistance (tens of kilohms). These two equivalent circuits require different designs when optimizing a readout circuit for low-noise performance. Although the low-noise circuits discussed in this paper are developed for terahertz imaging arrays, the design principles are general and can be applied to any voltage measurement in the presence of white and $1/f$ noise.[80, 81]

Section II of this paper discusses general noise theory as it applies to the analysis of the readout circuit, including quantification and reduction of detector thermal noise, and amplifier voltage and current noise in the low-frequency ($1/f$) and broadband regions of the spectrum. Section III describes a single amplifier channel, the basic building block of the readout electronics, including overall design considerations. This section concludes with an application of the noise theory developed in Section II to the specific readout circuit, and compares the theoretical noise performance with measured results, showing that it is important to identify constituent noise sources in the optimization of low-noise circuitry. Section IV describes two terahertz imaging array measurements, demonstrating the applicability of the readout electronics, and achieving detector limited noise performance.

A.3 Analysis of Noise Sources

The dominant noise sources in this work are generated by the detector and by the electronics, and can generally be divided into white noise and $1/f$ noise. These noise sources have unique frequency distributions and therefore useful quantities for describing noise are the *power* spectral density, $S(f)$, expressed in V^2/Hz, and the *voltage* spectral density, $S_v(f)^{1/2}$, expressed in $\mathrm{V}/\sqrt{\mathrm{Hz}}$. Throughout this work, single-sided spectral densities are used. The variance, σ_v^2, often called the noise power of a zero mean signal in V^2, is calculated from the single-sided noise power spectral density as[82]

$$\sigma_v^2 = \int_0^\infty S(f)\mathrm{d}f. \tag{A.1}$$

The general class of linear circuits (filters, gain blocks, etc.) are described by their amplitude transfer functions in the frequency domain, $H(\jmath 2\pi f)$. In combination with the noise power spectral density, $S(f)$, the transfer of noise power from the input of a linear circuit block to its output is expressed as[52]

$$\sigma_{v\mathrm{out}}^2 \quad = \quad \int_0^\infty S_{\mathrm{in}}(f)|H(\jmath 2\pi f)|^2\mathrm{d}f, \tag{A.2}$$

where $S_{\mathrm{in}}(f)$ includes the input-referred noise contribution of the network described by $H(\jmath 2\pi f)$.

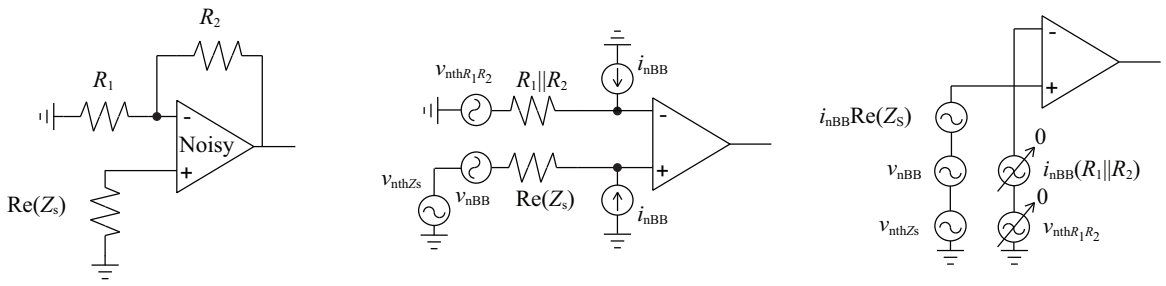
White noise is characterized by a uniform power density across frequency (neglecting the quantum correction [83]), and is most commonly generated by resistive material at temperatures above 0 K. Johnson[84] observed the lower limit on the mean-square voltage fluctuations of a measurement, due to a resistance R, with a noise bandwidth $f_2 - f_1$, at physical temperature T, is

$$\overline{V_{\mathrm{nth}}^2} = 4k_{\mathrm{B}}TR(f_2 - f_1). \tag{A.3}$$

The root-means-square (RMS) voltage can be expressed as the product of the thermal noise voltage spectral density, $S_{v\mathrm{nth}}^{1/2}$, and the square root of the equivalent noise bandwidth[85] (ENBW):

$$\sigma_{v\mathrm{nth}} = S_{v\mathrm{nth}}^{1/2}\sqrt{\mathrm{ENBW}}\,, \tag{A.4}$$

where $S_{v\mathrm{nth}}^{1/2} = \sqrt{4k_BTR}$. In the case of a complex impedance, R is replaced by the real part of



(a) Noisy components. (b) Equivalent voltage and current sources. (c) Equivalent voltage sources.

Figure A.2: A physical measurement is represented by noisy resistors and a noisy amplifier (a), or equivalently as noiseless components with explicit detector thermal noise ($v_{\mathrm{nth}Z\mathrm{s}}$), feedback thermal noise ($v_{\mathrm{nth}R1R2}$), op-amp voltage noise (v_{nBB}), and op-amp current noise (i_{nBB}) (b). Finally, noise sources can be represented as simplified equivalent voltage sources (c).

the complex impedance, Re($Z(f)$), and $S_{v\mathrm{nth}}^{1/2}$ becomes a function of frequency[80], such that

$$\sigma_{v\mathrm{nth}} = \sqrt{4k_B T R \int_0^\infty \mathrm{Re}(Z(f))|H(\jmath 2\pi f)|^2 \mathrm{d}f}\,. \tag{A.5}$$

In this work, a typical measurement consists of an AC-coupled detector with effective impedance, Z_s, a low-noise amplifier (LNA), and a feedback network, as shown in Fig. A.2(a). The three dominant sources of *white* noise in this measurement are thermal voltage noise (v_{nth}), and broadband amplifier voltage (v_{nBB}) and current (i_{nBB}) noise, as shown explicitly in Fig. A.2(b). For convenience, current noise is converted to voltage noise as shown in Fig. A.2(c).

The RMS voltage due to broadband voltage noise from the amplifier is calculated as

$$\sigma_{v\mathrm{nBB}} = S_{v\mathrm{nBB}}^{1/2}\sqrt{\mathrm{ENBW}}. \tag{A.6}$$

The RMS voltage due to broadband current noise from the amplifier depends on $Z_\mathrm{s}(f)$, and is calculated as

$$\sigma_{i\mathrm{nBB}} = \sqrt{\int_0^\infty (\mathrm{Re}(Z_\mathrm{s}(f))S_{i\mathrm{nBB}}^{1/2})^2|H(\jmath 2\pi f)|^2 \mathrm{d}f}\,. \tag{A.7}$$

The current noise through the feedback network, and the thermal noise of the feedback network can be neglected if they are small compared to the input circuit and detector, as shown in Fig. A.2(c).

The physical mechanisms behind $1/f$ noise are not well understood, but are universally characterized by a power spectral density proportional to $1/f$. In this work, we are concerned with flicker noise generated in a biased detector[86], and in the front-end of an amplifier.[76]

Uncorrelated noise power adds in quadrature,

$$\sigma_{v\text{tot}} = \sqrt{\sigma_{v1}^2 + \sigma_{v2}^2}\,, \tag{A.8}$$

so that one noise source will usually dominate the total measurement noise. The crossover point at which, for example, $1/f$ noise-dominance yields to white noise-dominance is called the $1/f$ knee, designated f_c in Fig. A.3.

A.3.1 Noise Reduction

The overall noise performance of a properly designed room-temperature imaging system is dominated by detector noise, not readout noise. Here, we explore techniques for reducing uncertainty in the final measurement due to noise in the readout electronics.

The source resistance of a room-temperature detector is optimized for best performance[78, 79, 87, 88] so, referring to (A.3), the primary method of thermal noise reduction in this work is the reduction of measurement bandwidth. In 1979, Grimbleby described the "Ideal Averaging Filter" [53] (IAF) as the most efficient method for white noise reduction in the measurement of a constant signal. The IAF integrates (filters) a noisy signal for a period of time, τ_{int}, during which the signal grows linearly and the noise grows as the square root of integration time. Thus, SNR increases as the square root of integration time. It is necessary that each IAF measurement be independent of previous measurements and the time-domain step-response be linear.

The "Gated Integrator" shown in the dashed box of Fig. A.4 approximates the ideal averaging filter.[82] The single-pole low-pass filter has an exponential integration response, however when $\tau_{\text{int}} << \text{RC}$, it is approximately linear. The gated integrator clears the previous signal, integrates for τ_{int}, and holds the signal for sampling and digitization. We refer to this white-noise reducing measurement technique as the "on" measurement protocol. The RMS voltage at the output of an "on" measurement is $\sigma = \sqrt{S_0/2\tau_{\text{int}}}$, in the presence of white noise only. However, in the presence of $1/f$ noise, fluctuations are divergent and a different technique is needed.

Figure A.3 shows the frequency domain voltage spectral density, $S_v(f)^{1/2}$, of a small signal, periodic at f_{mod}, in the presence of seemingly-overwhelming white and $1/f$ noise. The signal is

stronger than the white noise and $1/f$ noise at f_{mod}, so by applying appropriate filtering the modulated signal can be extracted. For $\sqrt{\tau_{\text{int}}}$ improvement in SNR, filter integration must be coherent and thus f_{mod} should be greater than f_{c}. This work attempts to measure a DC signal so it must be intentionally modulated. This type of measurement is referred to as phase sensitive detection[75] (PSD).

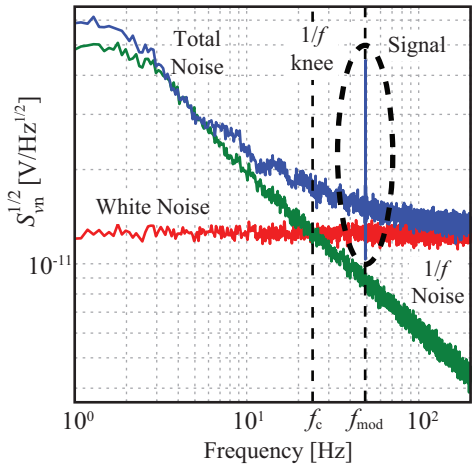


Figure A.3: A simulated low power signal at f_{mod} is buried in white and $1/f$ noise. Phase sensitive detection can extract the modulated signal from the noise with the required SNR, while low-pass filtering yields diminishing returns for increased integration time.

A realization of PSD is the "on-off" measurement protocol, depicted in Fig. A.4, consisting of two measurements taken every $1/f_{\text{mod}}$ seconds, synchronous with the modulated signal. The measurement in the first half-period, $\tau_{\text{int}} = 1/2f_{\text{mod}}$, is taken according to the "on" measurement protocol to reduce white noise. For the remaining half-period a similar measurement with the signal nulled is conducted, and referred to as the "off" measurement. The difference of the two measured voltages constitutes the "on-off" measurement protocol, which removes any $1/f$ drift below f_{mod}.

A.3.2 Measurement Protocol Comparison

The mean-square fluctuation of a sampled voltage, at the output of an analog circuit is calculated as[82]

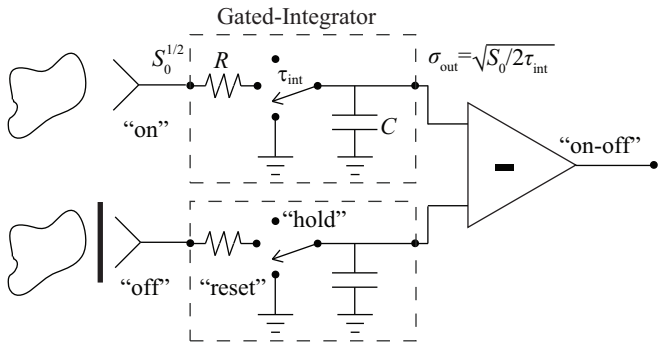


Figure A.4: The "on-off" measurement topology, for the reduction of white noise and $1/f$ noise is the difference of two subsequent "on" measurement protocols, one with the signal nulled. The gated-integrator is shown in the dashed box.

$$\sigma^2 = \int_0^\infty S(f)\,|W(\jmath 2\pi f)|^2\,\mathrm{d}f, \tag{A.9}$$

where $S(f)$ is the power spectrum of the input noise, and $W(\jmath 2\pi f)$ is the measurement protocol transfer function. Table A.1 summarizes the results of (A.9) for the "on" and "on-off" measurement protocols in the presence of white ($S = S_0$) and $1/f$ ($S = S_0 f_c/f$) noise. The white-noise variance of the "on" measurement (averaging for τ_{int}) is inversely proportional to integration time; however, the $1/f$-noise variance diverges with integration time. The variance of the "on-off" measurement also reduces white-noise variance (though variance is four times higher than the "on" measurement), and maintains the $1/f$-noise variance at a constant level. This is the preferred measurement protocol for this work. A similar technique was pioneered in the original Dicke radiometer [89], and is often used for bolometric radiometry measurements.[77]

A.4 Readout Single Channel Design

The readout electronics are designed to interface with an array of detectors so the complete readout is a parallelized array of digitally sampled channels. Each channel should accurately measure detector noise, which requires channel noise to be lower than detector noise. Section A.3 indicates that detector limited noise performance can be achieved through the proper application of a gated integrator employed in an "on-off" measurement protocol.

	σ^2 "on" $\left[\mathrm{V}^2\right]$	σ^2 "on-off" $\left[\mathrm{V}^2\right]$
$w(t)$	$\frac{1}{\tau}\mathtt{rect}\left(\frac{t-\frac{\tau}{2}}{\tau}\right)$	$\frac{2}{\tau}\mathtt{rect}\left(\frac{t-\frac{\tau}{4}}{\frac{\tau}{2}}\right) - \frac{2}{\tau}\mathtt{rect}\left(\frac{t-\frac{3\tau}{4}}{\frac{\tau}{2}}\right)$
$W(\jmath\omega)$	$e^{-\jmath\omega\frac{\tau}{2}}\dfrac{\sin(\omega\tau/2)}{\omega\tau/2}$	$2\jmath e^{-\jmath\omega\frac{\tau}{2}}\dfrac{\sin^2(\omega\tau/4)}{\omega\tau/4}$
White	$\dfrac{S_0}{2\tau}$ [53]	$\dfrac{2S_0}{\tau}$
$1/f$	divergent	$4\ln 2 S_0 f_c$ [82]

Table A.1: Output variation of an "on" measurement protocol diverges in the presence of $1/f$ noise, while an "on-off" measurement protocol maintains $1/f$ noise to a constant low level. $\omega = 2\pi f$, and τ is the integration time.

The front-end and back-end of such a measurement are shown in Fig. A.5(a), and Fig. A.5(b) respectively. The front-end amplifies the detector signal, and applies a variable ENBW limit ($1/2\tau_{\mathrm{int}}$) through the programmable gated integrator. The back-end buffers the gated integrator and performs additional amplification. A conceptual noise analysis provides further insight into the circuit parameters critical to detector limited performance (e.g., gain, ENBW).

If the equivalent input noise spectral density, $S_{v\mathrm{nin}}^{1/2}$, in Fig. A.5(a) is assumed constant over frequency, and the "on-off" measurement protocol is appropriately applied ($f_{\mathrm{mod}} > f_{\mathrm{c}}$), a simplified noise analysis need only consider constant white noise sources (Fig. A.2(b)).

The analysis proceeds as follows: (1) determine noise at the output of the gated integrator, σ_{RC}, due to the front-end (Fig. A.5(a)) as a function of the programmable integration time, τ_{int}, (2) determine noise at the output of the channel, σ_{out}, due to the back-end (Fig. A.5(b)), and (3) incoherently combine the two.

The noise at the output of the gated integrator due to front-end noise sources, according to (A.6), is

$$\sigma_{RC} = S_{v\mathrm{nin}}^{1/2} G_{\mathrm{in}} \frac{\tau_{\mathrm{int}}}{RC}\sqrt{\mathrm{ENBW}_{\mathrm{in}}}\,, \text{ or} \tag{A.10}$$

$$\sigma_{RC} = \frac{S_{v\mathrm{nin}}^{1/2} G_{\mathrm{in}}}{RC}\sqrt{\frac{\tau_{\mathrm{int}}}{2}}\,, \tag{A.11}$$

where $S_{v\mathrm{nin}}^{1/2}$ is the combined contribution of the voltage and current noise of the front-end amplifier.

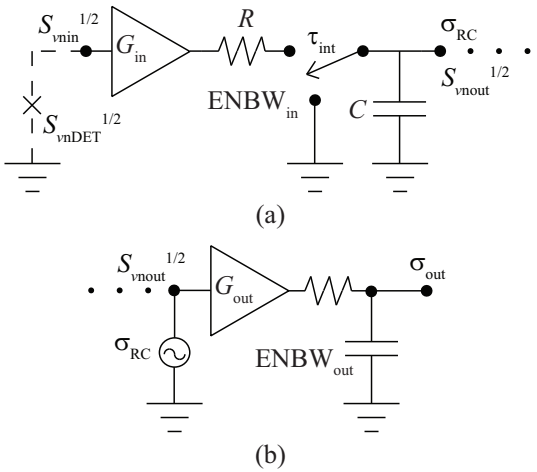


Figure A.5: The front-end of a single readout channel (a) amplifies and integrates the input signal and noise. The back-end of the channel provides buffering and additional gain (b).

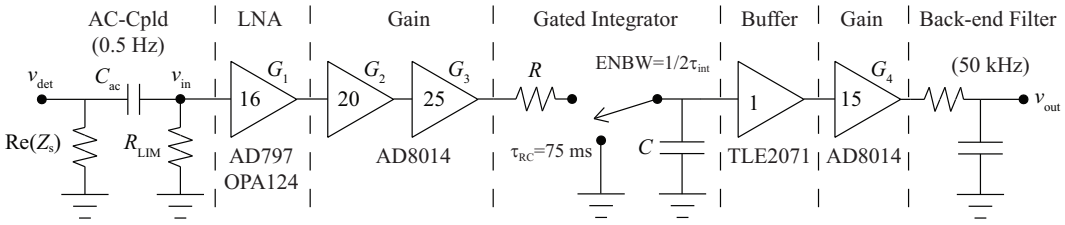


Figure A.6: A single low-noise electronic readout channel is shown in simplified schematic. Table A.2 summarizes the important specifications for each amplifier. The AD797/OPA124 is a low-noise amplifier, the AD8014 is a general purpose amplifier, and the TLE2071 is a low bias current buffer for the integrator.

From (A.8) the total readout channel noise is

$$\sigma_{\text{out}} = \sqrt{\underbrace{(\sigma_{RC} G_{\text{out}})^2}_{\text{Front-end}} + \underbrace{\left(S_{v\text{nout}}^{1/2} G_{\text{out}} \sqrt{\text{ENBW}_{\text{out}}}\right)^2}_{\text{Back-end}}}. \tag{A.12}$$

Similarly, the noise of the final digitized data is the incoherent addition of the output of the readout electronics, σ_{out}, and the digitizer noise floor, σ_{DAQ}:

$$\sigma_{\text{fin}} = \sqrt{\sigma_{\text{out}}^2 + \sigma_{\text{DAQ}}^2}\,. \tag{A.13}$$

As long as the first term in (A.13) dominates the second, the measurement will be dominated by the noise of the readout electronics and not the digitizer. If the first term in (A.12) dominates, (A.13) simplifies to

$$\sigma_{\text{fin}} \approx \frac{S_{v\text{nin}}^{1/2} G_{\text{in}} G_{\text{out}}}{RC} \sqrt{\frac{\tau_{\text{int}}}{2}}\,, \tag{A.14}$$

which is just the front-end noise amplified by G_{out}, as expected for front-end noise dominance. The noise spectral density of the front-end electronics and the detector are both limited by the same ENBW (i.e., of the gated integrator) so as long as $S_{v\text{nin}}^{1/2}$ is less than the detector noise spectrum, the entire measurement is detector limited.

The practical realization of the circuit in Fig. A.5 must take into account (1) the high input gain, G_{in}, required for low-level terahertz signals, and (2) the frequency dependence of the coupled detector impedance, $Z_{\text{s}}(f)$. The high gain (G_{in} up to $100,000$) is accomplished through multiple cascaded gain stages, each of which must be bandwidth limited so that they do not saturate by amplified front-end noise. The frequency dependence of $Z_{\text{s}}(f)$ is accounted for by integrating over frequency. The realized circuit is shown in Fig. A.6 with pertinent parameters listed in TAB. A.2.

Table A.2: OP-AMP Specifications

Spec	AD797	OPA124	AD8014	TLE2071
BW (MHz)	8	1.5	400	10
S_{vnBB} (nV/$\sqrt{\text{Hz}}$)	1.3	6	3.5	11-17
S_{inBB} (pA/$\sqrt{\text{Hz}}$)	3	0.0005	5	0.0028
$f_{c1/f}$ (Hz)	10	220	-	160
I_{bias} (nA)	250	0.001	5000	0.015
V_{off} (mV)	0.025	0.2	2.5	0.34-4

The predicted noise is shown in Fig. A.7 for the circuit with two different LNAs, the AD797 and the OPA124 (trade names are provided for technical clarity and do not imply endorsement by NIST). Detector thermal noise, σ_{vnthZs}, LNA voltage noise, σ_{vnBB}, and LNA current noise, σ_{inBB}, are shown as constituents of the total theoretical noise, σ_{fin}. The gray region, where σ_{vnthZs} is dominant, indicates detector limited noise performance. For $\text{Re}(Z_s) < 100$, constant voltage noise from both LNAs dominates noise performance. As $\text{Re}(Z_s)$ increases detector Johnson noise dominates the LNA voltage noise, first for the AD797, and then for the OPA124. For high $\text{Re}(Z_s)$ the AD797 current noise dominates, while the OPA124 remains in the detector limited regime, but with a roll-off due to the parallel R_{LIM} in the AC-coupling filter, as shown in Fig. A.6. The roll-off will shunt signal from the detector in larger proportion than the noise is reduced so R_{LIM} and C_{ac} should be adjusted for specific detector impedances, to be discussed in Section A.5.

In summary, this section has presented the design of a detector limited readout channel. The results useful for measurement applications are given in Fig. A.7 with the following conclusions:

(1) For $(100 < \text{Re}(Z_s) < 6\,\text{k}\Omega)$, such as bolometric detectors[79], the AD797 is detector limited and should be used because of its overall lower noise at low resistances.

(2) For $(\text{Re}(Z_s) < 2\,\text{k}\Omega)$, such as diode detectors[90, 78], the OPA124 is detector limited; but, the input circuit should be set to prevent shunted signal.

A.5 Example Terahertz Imaging Measurements

The detector limited readout channel described in the previous section is parallelized for terahertz imaging array measurements. Figure A.1 shows the measurement procedure starting with a

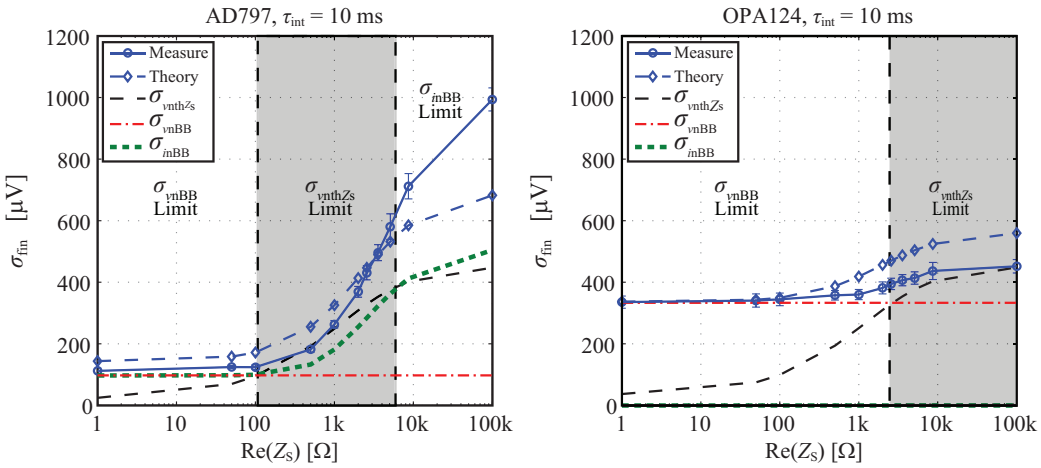


Figure A.7: Detector thermal noise ($\sigma_{v\mathrm{nth}Z\mathrm{s}}$), amplifier voltage ($\sigma_{v\mathrm{nBB}}$) and current noise ($\sigma_{i\mathrm{nBB}}$) add in quadrature for the total theoretical noise (dashed, diamond) showing good agreement with measured total noise (solid, circle). The gray regions indicate ideal operating regions in the "$\sigma_{v\mathrm{nth}Z\mathrm{s}}$ Limit". The AD797 is ideal for low resistance detectors and the OPA124 for high resistance detectors.

mechanical chopper which provides the modulation (f_{mod}) of blackbody radiation. A video-rate of 30 fps allows each channel a maximum integration time of $\tau_{\mathrm{int}} = 1/(2\times 30\,\mathrm{fps}) = 16.5\,\mathrm{ms/frame}$ for the "on" and the "off" half-period. The digitizer records each of the channel voltages sequentially, once per "on" and "off" state. Finally, the CPU performs the PSD algorithm. Additional averaging or other signal processing can be performed at the expense of frame update rate.

Each element in the detector array is an antenna-coupled power detector, characterized by dynamic resistance, R_{dyn}, and intrinsic current responsivity, $\mathfrak{R}_I$, which are calculated from a Taylor expansion of the DC IV-curves as[88, 91]

$$R_{\mathrm{dyn}} = \frac{\mathrm{d}V}{\mathrm{d}I}\ [\Omega], \text{ and} \tag{A.15}$$

$$\mathfrak{R}_I = \frac{\mathrm{d}^2 I/\mathrm{d}V^2}{2\mathrm{d}I/\mathrm{d}V}\ [\mathrm{A/W}]. \tag{A.16}$$

The zero-bias dynamic resistance, R_{zb}, is used to predict detector white noise, while responsivity is a measure of the signal from the device. The intrinsic voltage responsivity is calculated as

$$\mathfrak{R}_V = \mathfrak{R}_I R_{\mathrm{zb}}\ [\mathrm{V/W}]. \tag{A.17}$$

A total system thermal responsivity can be calculated from measured parameters as

$$\mathfrak{R}_T = \frac{\Delta V_{\mathrm{out}}/G_{\mathrm{tot}}}{\Delta T_{\mathrm{bb}}} = \frac{\Delta V_{\mathrm{det}}}{\Delta T_{\mathrm{bb}}}\ [\mathrm{V/K}], \tag{A.18}$$

where ΔT_{bb} is the radiometric temperature difference between two blackbody targets, G_{tot} is the readout gain, ΔV_{out} is the measured voltage difference at the output of the readout due to ΔT_{bb}, and ΔV_{det} is the measured voltage difference referred to the detector. This measurement includes coupling, mismatch, and other efficiencies between an ideal blackbody and the output of the readout.

A standard measure of performance which includes responsivity and noise, is the noise equivalent temperature difference (NETD), usually normalized to video rate (30 Hz):

$$\mathrm{NETD} = \frac{\mathrm{NEP}}{k_{\mathrm{B}}\Delta f}\sqrt{\frac{1}{2\times 33\,\mathrm{ms}}}\ [\mathrm{K}]. \tag{A.19}$$

NETD can be calculated directly from the measured, input-referred noise (σ_{in}) and thermal

Table A.3: System 1 array characterization from DC IV-curves provides zero-bias resistance (R_{zb}), current responsivity ($\mathfrak{R}_I$), and voltage responsivity ($\mathfrak{R}_V$) from (A.15)-(A.17). Equation (A.20) gives antenna mismatch responsivity ($\mathfrak{R}_V(1-|S_{11}|^2)$), thermal responsivity ($\mathfrak{R}_T$), and NETD for an RF bandwidth of 10 GHz and a 33 ms post-detection integration time.

Diode	1	2	3	4	5	6	7	8
R_{zb} [kΩ]	-	12.1	21	-	28.6	20.3	20.4	-
$\mathfrak{R}_I$ [A/W]	-	5.7	14.5	-	13.5	14.7	14.7	-
$\mathfrak{R}_V$ [kV/W]	-	69	300	-	390	300	300	-
$\mathfrak{R}_V(1-\lvert S_{11}\rvert^2)$ [kV/W]	-	8.5	22	-	21	23	23	-
$\mathfrak{R}_T$ [nV/K]	-	1.2	3	-	2.9	3.1	3.1	-
$\mathrm{NETD}_{33\,\mathrm{ms},10\,\mathrm{GHz}}$ [K]	-	47	23	-	29	23	23	-

responsivity ($\mathfrak{R}_T$)

$$\mathrm{NETD} = \frac{S_v^{1/2}\sqrt{\frac{1}{2\times 0.033}}}{\mathfrak{R}_T} = \frac{\sigma_{\mathrm{in}}}{\mathfrak{R}_T}, \tag{A.20}$$

where σ_{in} and $\mathfrak{R}_T$ are referred to a 33 ms integration time.

Measurements with the readout circuit from Fig. A.6 were performed for two terahertz imaging systems: (1) an 8-element, horn-antenna coupled diode array similar to Ref. [87] available at NIST-Boulder, and (2) a 32-element array of horn-antenna coupled diodes located at the Army Research Laboratory (ARL) in Adelphi, MD.[92]

A.5.1 System 1: High $\mathrm{Re}(Z_{\mathrm{DET}})$

Figure A.8 shows a photograph of the 8-element terahertz imaging array at NIST-Boulder employing ErAs/InAlGaAs diode detectors.[87] The diodes were initially characterized with DC IV-curve measurements for noise and responsivity predictions. As a result of multiple fabrication runs, the diodes exhibit significant non-uniformity. At the time of these measurements diodes 1, 4, and 8 were no longer functional, and diode 2 had poor performance. Neglecting these diodes, the typical zero-bias resistance is 20 kΩ, with voltage responsivity of approximately 3 nV/K, yielding an expected NETD between 20 and 30 K. Table A.3 summarizes the results of the IV-curve characterization.

The readout was originally configured for use with a 500 Ω bolometric detector[79] so the AC-coupling input filter included a 4.2 kΩ shunt resistor (R_{LIM}). Using this default configuration, the measured input-referred noise (with an OPA124) for a typical diode, with a measurement gain

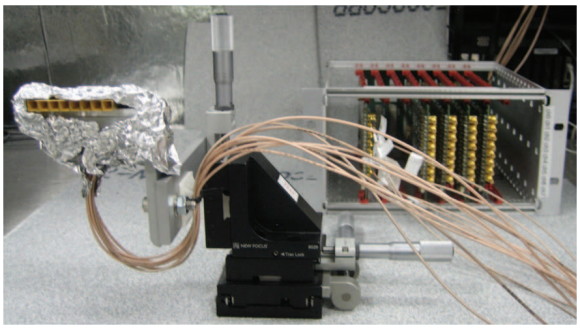


Figure A.8: An 8-element antenna-coupled, zero-bias diode imaging array is shielded to reduce interference. A chassis hosts up to eight, 8-channel readout boards. Note the shielding around the array to reduce interfering signals.

of 14,750, is approximately $510\,\mu\text{V}/14{,}750 = 35\,\text{nV}$, as shown for diode 5 in Fig. A.9. In a 10 ms integration time, that is $35\,\text{nV}/\sqrt{(1/(2\times 10\,\text{ms})} = 5\,\text{nV}/\sqrt{\text{Hz}}$, which is close to the $5.7\,\text{nV}/\sqrt{\text{Hz}}$ expected for an effective resistance of $2\,\text{k}\Omega$ ($20\,\text{k}\Omega$ detector in parallel with the input circuit).

To measure $\mathfrak{R}_T$ directly, a mechanical chopper is placed in front of the diode detector array. Quadrants of the chopper wheel present alternate hot (295 K Eccosorb®) and cold (77 K liquid nitrogen) loads to the diode array. The input referred response is typically $220\,\mu\text{V}/14{,}750 = 14.9\,\text{nV}$, and maximally $277\,\mu\text{V}/14,750 = 18.8\,\text{nV}$, as shown in Fig. A.10. From (A.18), the input-referred thermal responsivity is $\mathfrak{R}_T = 18.8\,\text{nV}/218\,\text{K} = 86\,\text{pV/K}$. The NETD, from (A.20) is 226 K which is significantly higher than expected. Given the practical considerations discussed in Section A.4 of interfacing with a specific detector resistance, if the $4.2\,\text{k}\Omega$ shunt resistor in the input circuit is replaced with a larger resistor (e.g., $1\,\text{M}\Omega$[1]), the predicted coupled responsivity increases by a factor of 10.5. Remeasuring noise with this input impedance, and including an optical coupling efficiency of approximately 30%, as discussed in detail in Ref. [92], we obtain the expected 29 K NETD from TAB. A.3, which confirms that the increased NETD is completely attributed to shunted signal when using $R_{\text{LIM}} = 4.2\text{k}\Omega$, and not excess noise. Thus, the recommended configuration for optimal noise, and responsivity for these diodes is $R_{\text{LIM}} \approx 1\text{M}\Omega$ with the OPA124.

[1] The extremely low bias current of the OPA124 allows for very large R_{LIM} without saturating the circuit.

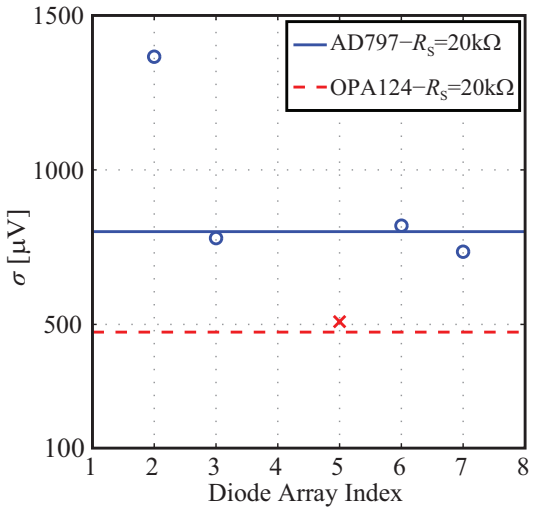


Figure A.9: Measured noise for diodes 3, 6, and 7 agrees with theory with an AD797 LNA. Diode 5 agrees with theory with an OPA124. Horizontal lines indicate the measured noise with a 20 kΩ shielded resistor. Note that diode 2 has anomalously high noise (as expected due to low resistance and responsivity from IV-curves), and diodes 1, 4, and 8 were damaged.

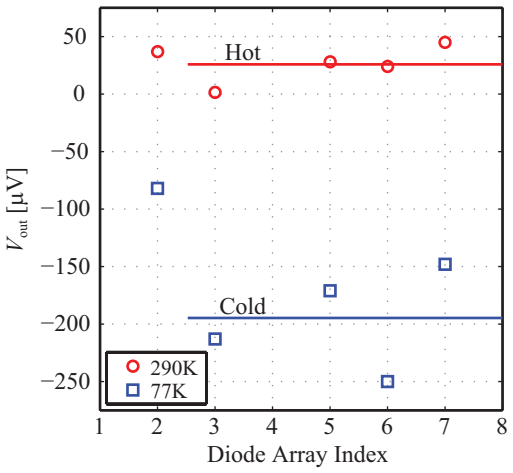


Figure A.10: The typical response of each diode from the "hot" to "cold" state, at the output of the readout circuit, in a 10 ms integration is approximately 220 μV. Diode 2 has anomalously low signal and diodes 1, 4, and 8 were damaged.

A.5.2 System 2: Low Re(Z_{DET})

The readout circuit described in this work was also sent to ARL for independent measurement with a 32 element array of Sb-based backward tunnel diodes.[92] The Re(Z_{det}) was 2 kΩ[78], much lower than for the system measured at NIST-Boulder. R_{LIM} was set to 6.8 kΩ to maximize the signal, and the AD797 was used as the LNA. Measured NETDs across the array of diodes were reported to be typically 40 K, with ranges from 20-80 K which were deemed too high to be useful by a factor of four.[92] However, it was noted that degraded NETD was primarily due to excessive interference which is reduced in system 1 by shielding as seen in Fig. A.8. Using the responsivity of 2.2 nV/K from Ref. [92], and the noise performance of a shielded 2 kΩ detector of 5 nV/$\sqrt{\mathrm{Hz}}$ (as demonstrated in Fig. A.7), we expect an NETD of 8.8 K. However, without proper shielding and careful grounding this NETD rises significantly, underscoring the importance of shielding in order to achieve the theoretical detector limited performance of the readout circuit.

A.6 Conclusion

The measurements from Section A.5 show the applicability of the readout circuit to various imaging systems. Correctly configured, this circuit can achieve detector limited performance with any direct detector at THz, or other frequencies. The recommended readout configuration is determined from detector *IV*-curves, as shown in Fig. A.11, and is summarized in TAB. A.4 for various systems from the literature. Zero-bias dynamic resistance, R_{zb}, motivates selection of the LNA based on the detector limited regions defined in Fig. A.7. R_{LIM} is selected to simultaneously prevent the LNA bias current from saturating the circuit, and from shunting too much of the detector signal to ground, while still enforcing a low-frequency cutoff of approximately 0.5 Hz. Figure A.11 shows the published *IV*-curves for various detectors in the literature [79, 93, 87, 94], with R_{zb} ranging from 500 Ω to almost 1 MΩ. The recommended readout configuration for each detector is given in Tab. A.4.

The main contribution of this paper is a flexible readout circuit design, made with off-the-shelf

Table A.4: Zero-Bias Resistance ($R_{\rm zb}$), input circuit shunt resistance ($R_{\rm LIM}$), and LNA selection for each detector determine an expected Responsivity ($\mathfrak{R}_T$), and NETD.

	Recommended Config.			Measurements*			Limit
	$R_{\rm zb}$ [Ω]	LNA	$R_{\rm lim}$ [Ω]	$S_{vn}^{1/2}$ $[\frac{\rm nV}{\sqrt{\rm Hz}}]$	$\mathfrak{R}_T$ $[\frac{\rm nV}{\rm K}]$	NETD	NETD
Sys 1 - A.5.1	28 k	OP124	4.2 k	5	.086	226 K	29 K
Sys 1 - $R_{\rm in} = 1\,{\rm M}\Omega$	28 k	OP124	1 M	21	0.9	89 K	29 K
Sys 2 - A.5.2[92]	2 k	AD797	6.8 k	22	2.2	40 K	2.2 K
Sys 2 - shielded	2 k	AD797	6.8 k	5	2.2	8.8 K	2.2 K
Nb bolo [79]	505	AD797	4.2 k	*2.9*	*0.7*[i]	*16 K*	-
Hetero diode [93]	860	AD797	4.2 k	*2.4*[ii]	*0.6*[ii]	*24 K*	11 K
Hetero diode [90]	6.5 k	OP124	1 M	*10.3*	*2.3*[iii]	*17 K*	-
Sb diode [94]	18 k	OP124	1 M	*17*	*3.9*[iv]	*17 K*	-
Hetero diode [90]	20 k	OP124	1 M	*18*	*2.5*[v]	*28 K*	-
Hetero diode [87]	955 k	OP124	2 M	*125*	*3.5*[vi]	*9.7 K*	2.47 K

*Results are based upon measurements in this work for the two systems discussed, while italicized entries are calculated for various systems based upon results in the literature.
[i] Ref. [79] shows $\beta = 53\,{\rm V/W}$ at $I_{\rm bias} = 200\,\mu{\rm A}$. $\Delta f_{\rm RF} = 1\,{\rm THz}$ [70].
[ii] Ref. [93] shows NEP=$2\,{\rm pW}/\sqrt{\rm Hz}$, and $\mathfrak{R}_V = 1200\,{\rm V/W}$, and assuming $\Delta f_{\rm RF} = 35\,{\rm GHz}$.
[iii] Ref. [90] shows $\mathfrak{R}_I = 11.7\,{\rm A/W}$. The horn antenna has approximately $\Delta f_{\rm RF} = 10\,{\rm GHz}$, with an impedance of 400 Ω. So, after mismatch loss, $\mathfrak{R}_T = 2.3\,{\rm nV/K}$.
[iv] Ref. [94] shows $\mathfrak{R}_V = 3687\,{\rm V/W}$, including 50 Ω mismatch. So, in the same 400 Ω antenna as Ref. [87], $\mathfrak{R}_V = 333,000\,{\rm V/W}$,or $\mathfrak{R}_T = 3.9\,{\rm nV/K}$, if $\Delta f_{\rm RF} = 10\,{\rm GHz}$.
[v] Ref. [90] shows $\mathfrak{R}_I = 11.7\,{\rm A/W}$. The horn antenna has approximately $\Delta f_{\rm RF} = 10\,{\rm GHz}$, with an impedance of 400 Ω. So, after mismatch loss, $\mathfrak{R}_T = 2.5\,{\rm nV/K}$.
[vi] Figure 6 of ref. [87] shows $\mathfrak{R}_T = 50\,{\rm nV/K}$, so assuming thermal noise of detector is dominant, NETD=9.7 K.

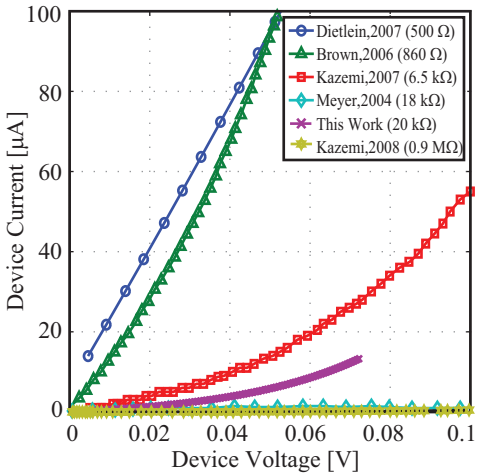


Figure A.11: Device *IV*-curves are used to predict zero-bias dynamic resistance which determines how the readout input circuit should be configured (R_{LIM}, and LNA). Devices references are as noted for trace markers: circle[79], triangle[93], square[90], diamond[94], cross$^{(\mathrm{This\ Work})}$, hexagram[87].

components that can be easily adapted to any direct detector for achieving detector limited noise performance for video rate arrays. Recommendations are provided for maximizing the coupled responsivity of any diode based on a DC *IV*-curve characterization. The approach is demonstrated on arrays of up to 64 detectors, but is scalable to imaging systems with much larger arrays.

Acronyms and Abbreviations

AFM atomic force microscope

AC alternating current

CAT Circuit Analysis Tools

DUT device under test

DC direct current

ENBW equivalent noise bandwidth

FOV field of view

FEM finite element method

FEMM Finite Elements Method for Magnetics

GaAs gallium-arsenide

HFSS Ansys HFSS^{TM}

IRIS Integrity and Reliability of Integrated Circuits

IC integrated circuit

IMD intermodulation distortion

IAF ideal averaging filter

LVNP Lock-in Vector Near-field Probe

LPF low-pass filter

MMIC monolithic microwave integrated circuit

MWO Applied Wave Research Microwave Office

MATLAB MathWorks MATLAB®

MTF modulation transfer function

NFMM near-field microwave microscope

NFMM near-field microwave microscopy

NFMD near-field microwave detection

NSOM near-field scanning optical microscope

NSOM near-field scanning optical microscopy

ODE ordinary differential equation

PSD phase sensitive detector

PSD phase sensitive detection

PDE partial differential equation

PEC perfect electric conductor

PCB printed circuit board

Q quality factor

Q-factor quality factor

Q_L loaded quality factor

Q_U unloaded quality factor

Q_E external quality factor

SNR signal-to-noise ratio

SPM scanning probe microscope

STM scanning tunneling microscope

SEM scanning electron microscopy

SOL short-open-load

TRUST-IC TRUST in Integrated Circuits

THz terahertz

VNA vector network analyzer

Bibliography

[1] R. Bracewell, *The Fourier transform and its applications*. McGraw-Hill series in electrical and computer engineering, McGraw Hill, 2000. 2

[2] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, pp. 429–441, Nov. 1946. 2

[3] A. M. H. Wong and G. V. Eleftheriades, "Temporal pulse compression beyond the fourier transform limit," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, pp. 2173–2179, Sept. 2011. 2

[4] E. Abbe, "Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung," *Archiv für Mikroskopische Anatomie*, vol. 9, pp. 413–418, Dec. 1873. 2, 5

[5] C. Bryant and J. Gunn, "Noncontact technique for the local measurement of semiconductor ResistivitY," *The Review of Scientific Instruments*, vol. 36, pp. 1614–1617, Nov. 1965. 2, 9

[6] E. A. Ash and G. Nicholls, "Super-resolution aperture scanning microscope," *Nature*, vol. 237, pp. 510–512, June 1972. 2, 10

[7] C. Gao, T. Wei, F. Duewer, Y. Lu, and X. D. Xiang, "High spatial resolution quantitative microwave impedance microscopy by a scanning tip microwave near-field microscope," *Applied Physics Letters*, vol. 71, pp. 1872–1874, Sept. 1997. 2, 3, 75

[8] M. Tabib-Azar, D. Su, A. Pohar, S. R. LeClair, and G. Ponchak, "0.4 μm spatial resolution with 1 GHz ($\lambda = 30\,\mathrm{cm}$) evanescent microwave probe," *Review of Scientific Instruments*, vol. 70, pp. 1725–1729, Mar. 1999. 2, 8, 11

[9] L. Zhang, Y. Ju, A. Hosoi, and A. Fujimoto, "Microwave atomic force microscopy imaging for nanometer-scale electrical property characterization," *Review of Scientific Instruments*, vol. 81, no. 12, p. 123708, 2010. 2

[10] X. Wu and O. M. Ramahi, "Near-field scanning microwave microscopy for detection of subsurface biological anomalies," in *IEEE Antennas and Propagation Society International Symposium, 2004*, vol. 3, pp. 2444– 2447 Vol.3, IEEE, June 2004. 2

[11] N. Okazaki, H. Odagawa, Y. Cho, T. Nagamura, D. Komiyama, T. Koida, H. Minami, P. Ahmet, T. Fukumura, Y. Matsumoto, M. Kawasaki, T. Chikyow, H. Koinuma, and T. Hasegawa, "Development of scanning microwave microscope with a lumped-constant resonator probe for high-throughput characterization of combinatorial dielectric materials," *Applied Surface Science*, vol. 189, pp. 222–226, 2002. 2, 3, 67, 68

[12] W. Kundhikanjana, K. Lai, M. A. Kelly, and Z. Shen, "Cryogenic microwave imaging of metalâĂŞinsulator transition in doped silicon," *Review of Scientific Instruments*, vol. 82, no. 3, p. 033705, 2011. 2, 3

[13] "High performance microchip supply," Tech. Rep. ADA435563, Defense Science Board - Office of the Undersecretary of Defense For Acquisition, Technology, and Logistics, Feb. 2005. 3

[14] "Trust for integrated circuits (darpa-baa-06-40)," Tech. Rep. DARPA-BAA-06-40, Defense Advanced Research Projects Agency, May 2006. 3

[15] "Integrity and reliability in integrated ciruits (darpa-baa-10-33)," Tech. Rep. DARPA-BAA-10-33, Defense Advanced Research Projects Agency, 2010. 3

[16] "Circuit analysis tools (iarpa-baa-09-09)," Tech. Rep. IARPA-BAA-09-09, Intelligence Advanced Research Projects Activity - Office of the Director of National Intelligence, 2009. 3

[17] K. Lee, H. Melikyan, A. Babajanyan, T. Sargsyan, J. Kim, S. Kim, and B. Friedman, "Visualization of magnetic domains by near-field scanning microwave microscope," *Ultramicroscopy*, vol. 109, pp. 889–893, July 2009. PMID: 19342175. 3

[18] K. Lai, M. Nakamura, W. Kundhikanjana, M. Kawasaki, Y. Tokura, M. A. Kelly, and Z. Shen, "Mesoscopic percolating resistance network in a strained manganite thin film," *Science*, vol. 329, pp. 190 –193, July 2010. 3

[19] S. Huang, H. M. Christen, and M. E. Reeves, "Parameter-free extraction of Thin-

Film dielectric constants from scanning near field microwave microscope measurements," *arXiv:0909.3579*, Sept. 2009. 3, 32

[20] A. Imtiaz, T. M. Wallis, S. H. Lim, J. Chisum, Z. Popović, and P. Kabos, "Near-field antenna as a scanning microwave probe for characterization of materials and devices," in *Antennas and Propagation (EuCAP), 2010 Proceedings of the Fourth European Conference on*, pp. 1–3, 2010. 3, 80

[21] J. D. Chisum, M. Ramirez-Velez, and Z. Popović, "Planar circuits for non-contact Near-Field microwave probing," in *IEEE MTT EuMW Digest*, 2009. 3, 8, 43, 84, 85, 92

[22] Z. Ren, M. Boybay, and O. Ramahi, "Near-Field probes for subsurface detection using Split-Ring resonators," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 59, no. 2, pp. 488–495, 2011. 3

[23] C. Plassard, E. Bourillot, J. Rossignol, Y. Lacroute, E. Lepleux, L. Pacheco, and E. Lesniewska, "Detection of defects buried in metallic samples by scanning microwave microscopy," *Physical Review B*, vol. 83, p. 121409, Mar. 2011. 3, 85

[24] S. C. Cripps, "Probing times [Microwave bytes]," *IEEE Microwave Magazine*, vol. 10, pp. 28–34, Feb. 2009. 3

[25] H. P. Huber, M. Moertelmaier, T. M. Wallis, C. J. Chiang, M. Hochleitner, A. Imtiaz, Y. J. Oh, K. Schilcher, M. Dieudonne, J. Smoliner, P. Hinterdorfer, S. J. Rosner, H. Tanbakuchi, P. Kabos, and F. Kienberger, "Calibrated nanoscale capacitance measurements using a scanning microwave microscope," *Review of Scientific Instruments*, vol. 81, no. 11, p. 113701, 2010. 4, 64, 67

[26] C. Gao and X. Xiang, "Quantitative microwave near-field microscopy of dielectric properties," *Review of Scientific Instruments*, vol. 69, pp. 3846–3851, Nov. 1998. 4, 67

[27] G. Vaschenko, C. Brewer, F. Brizuela, Y. Wang, M. A. Larotonda, B. M. Luther, M. C. Marconi, J. J. Rocca, C. S. Menoni, E. H. Anderson, W. Chao, B. D. Harteneck, J. A. Liddle, Y. Liu, and D. T. Attwood, "Sub-38 nm resolution tabletop microscopy with 13 nm wavelength laser light," *Optics Letters*, vol. 31, pp. 1214–1216, May 2006. 5

[28] E. H. Synge, "A suggested method for extending microscopic resolution into the ultra-microscopic region," *Philosophical Magazine Series 7*, vol. 6, no. 35, p. 356, 1928. 5, 6

[29] S. Wang, "X-ray microtomography tools for advanced IC packaging failure analysis," in

Microelectronic Failure Analysis Desk Reference, The Electronic Device Failure Society Desk Reference Committee, pp. 260–267, ASM International, 5 ed., 2004. 6

[30] "12th national toxicology program," tech. rep., Department of Health and Human Services, 2011. 6

[31] M. Berry, E. Wolf, N. Bloembergen, N. Erez, and D. Greenberger, *Progress in Optics*, pp. 145–148. No. v. 50 in Progress in Optics, Elsevier, 2007. 6

[32] E. H. Synge, "An application of piezoelectricity to microscopy," *Philosophical Magazine*, vol. 13, p. 297, 1932. 7

[33] H. A. Bethe, "Theory of diffraction by small holes," *Physical Review*, vol. 66, p. 163, Oct. 1944. 7

[34] J. Howard H. Pattee, "The scanning X-Ray microscope," *Journal of the Optical Society of America*, vol. 43, pp. 61–62, Jan. 1953. 8

[35] J. A. O'Keefe, "Resolving power of visible light," *Journal of the Optical Society of America*, vol. 46, p. 359, May 1956. 9

[36] A. V. Baez, "Is resolving power independent of wavelength possible? an experiment with a sonic "macroscope"," *Journal of the Optical Society of America*, vol. 46, p. 901, Oct. 1956. 9

[37] G. A. Massey, "Microscopy and pattern generation with scanned evanescent waves," *Applied Optics*, vol. 23, no. 5, pp. 658–660, 1984. 11, 15

[38] J. R. Matey and J. Blanc, "Scanning capacitance microscopy," *Journal of Applied Physics*, vol. 57, pp. 1437–1444, Mar. 1985. 11

[39] B. R. Epstein, M. A. Gealt, and K. R. Roster, "The use of coaxial probes for precise dielectric measurments: A re-evaluation," in *Microwave Symposium Digest, MTT-S International*, vol. 87, pp. 255–258, MTT, June 1987. 11

[40] R. J. Gutmann, J. M. Borrego, P. Chakrabarti, and M. Wang, "Microwave scanning microscopy for planar structure diagnostics," in *Microwave Symposium Digest, MTT-S International*, vol. 87, pp. 281–284, MTT, June 1987. 11

[41] M. Tabib-Azar, neil S Shoemaker, and S. Harris, "Non-destructive characterization of

materials by evanescent microwaves," *Measurement Science and Techhnology*, vol. 4, pp. 583–590, May 1993. 11

[42] Nanosurf, *The Nanosurf Easyscan 2 AFM Specifications*, 2011. 12, 75

[43] E. F. Kuester and D. C. Chang, *Theory of Waveguides and Transmission Lines*, pp. 193–194. Boulder, CO: Course Notes, 1998. 15

[44] R. E. Collin, *Antennas and Radiowave Propagation*, ch. 1. New York: McGraw-Hill, 1985. 19, 32

[45] A. Imtiaz and S. M. Anlage, "Effect of tip geometry on contrast and spatial resolution of the near-field microwave microscope," *Journal of Applied Physics*, vol. 100, pp. 044304–1 to 044304–8, 2006. 24, 75, 80

[46] "Agilent impedance measurement handbook: A guide to measurement technology and techniques," Tech. Rep. 5950-3000, Agilent Technologies, June 2009. 24, 27

[47] J. Bray and L. Roy, "Measuring the unloaded, loaded, and external quality factors of one- and two-port resonators using scattering-parameter magnitudes at fractional power levels," *Microwaves, Antennas and Propagation, IEE Proceedings* -, vol. 151, no. 4, pp. 345–350, 2004. 25

[48] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th Edition*. Mineola, N.Y.: Dover Publications, 1972. 34

[49] G. Mesa, E. Dobado-Fuentes, and J. J. Saenz, "Image charge method for electrostatic calculations in field-emission diodes," *Journal of Applied Physics*, vol. 79, no. 1, p. 39, 1996. 34

[50] D. Cheng, *Field and Wave Electromagnetics, 2nd Edition*, pp. 172–174. Addison-Wesley, 1989. 35

[51] T. Wei, X. D. Xiang, W. G. Wallace-Freeman, and P. G. Schultz, "Scanning tip microwave near-field microscope," *Applied Physics Letters*, vol. 68, pp. 3506–3508, June 1996. 51, 67

[52] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*. New York, NY: IEEE Press, 1987. 55, 99

[53] J. B. Grimbleby, "The ideal averaging filter: its applications and realizations," *The Radio and Electronic Engineer*, vol. 49, p. 530–534, Oct. 1979. 56, 101, 104

[54] K. Iwaya, R. Shimizu, T. Hashizume, and T. Hitosugi, "Systematic analyses of vibration noise of a vibration isolation system for high-resolution scanning tunneling microscopes," *Review of Scientific Instruments*, vol. 82, no. 8, p. 083702, 2011. 60

[55] S. F. Adam, "A new precision automatic microwave measurement system," *IEEE Transactions on Instrumentation and Measurement*, vol. 17, pp. 308–313, Dec. 1968. 63

[56] H. J. Eul and B. Schiek, "A generalized theory and new calibration procedures for network analyzer self-calibration," *IEEE Transactions on Microwave Theory and Techniques*, vol. 39, pp. 724–731, Apr. 1991. 64

[57] T. Y. Otaoshi, "Improved RF calibration techniques: Scattering matrix of a reduced multiport," Tech. Rep. 37-56 Vol. II, NASA JPL, Pasadena, CA, Mar. 1969. 64

[58] J. D. Chisum and Z. Popović, "A scanning lock-in vector near-field probe for noise limited microwave measurements," July 2010. 65, 80, 84

[59] J. D. Chisum, E. N. Grossman, and Z. Popović, "A general approach to low noise readout of terahertz imaging arrays," *Review of Scientific Instruments*, vol. 82, no. 6, p. 065106, 2011. 65, 84, 85

[60] J. D. Chisum and Z. Popović, "Under review: Performance limitations and measurement analysis of a near-field microwave microscope for non-destructive and sub-surface detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. V, p. P, sept 2011. 65, 80, 84

[61] K. Lai, W. Kundhikanjana, M. Kelly, and Z. X. Shen, "Modeling and characterization of a cantilever-based near-field scanning microwave impedance microscope," *Review of Scientific Instruments*, vol. 79, no. 6, p. 063703, 2008. 67

[62] A. Imtiaz, T. Baldwin, H. T. Nembach, T. M. Wallis, and P. Kabos, "Near-field microwave microscope measurements to characterize bulk material properties," *Applied Physics Letters*, vol. 90, p. 243105, June 2007. 68

[63] X. Zhang, X. Wang, F. Xu, Y. Ma, and C. K. Ong, "High frequency dielectric properties distribution of $BiFeO_3$ thin film using near-field microwave microscopy," *Review of Scientific Instruments*, vol. 80, pp. 114701–4, Nov. 2009. 68

[64] K. Lai, W. Kundhikanjana, M. A. Kelly, and Z. X. Shen, "Calibration of shielded microwave probes using bulk dielectrics," *Applied Physics Letters*, vol. 93, no. 12, p. 123105, 2008. 74

[65] A. Imtiaz and S. M. Anlage, "A novel STM-assisted microwave microscope with capacitance and loss imaging capability," *Ultramicroscopy*, vol. 94, pp. 209–216, Apr. 2003. 75

[66] G. D. Boreman, *Modulation Transfer Function in Optical and Electro-optical Systems*, vol. TT52 of *Tutorial Texts in Optical Engineering*. Bellingham, WA: SPIE Press, 2001. 80

[67] S. K. Cheung, T. P. Halloran, W. H. Weedon, and C. P. Caldwell, "MMIC-Based quadrature hybrid Quasi-Circulators for simultaneous transmit and receive," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 489–497, Mar. 2010. 85

[68] K. J. Vanhille, D. L. Fontaine, C. Nichols, D. S. Filipovic, and Z. Popović, "Quasi-planar high-Q millimeter-wave resonators," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 2439–2446, June 2006. 85, 91

[69] C. Chi and G. M. Rebeiz, "Conductor-loss limited stripline resonator and filters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, pp. 626–630, Apr. 1996. 85, 91

[70] E. N. Grossman, C. R. Dietlein, J. Chisum, A. Luukanen, J. E. Bjarnasson, and E. R. Brown, "Spectral decomposition of ultra-wide-band terahertz imagery," in *Passive Millimeter-Wave Imaging Technology X* (R. Appleby and D. A. Wikner, eds.), vol. 6548, (Orlando, FL, USA), pp. 654807–8, SPIE, Apr. 2007. 97, 114

[71] R. Appleby, P. Coward, and J. N. Sanders-Reed, "Evaluation of a passive millimeter-wave (PMMW) imager for wire detection in degraded visual conditions," in *Proceedings of SPIE*, (Orlando, FL, USA), pp. 73090A–73090A–8, 2009. 97

[72] P. Siegel, "Terahertz technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 910–928, 2002. 97

[73] A. Luukanen, L. Gronberg, P. Helisto, J. S. Penttila, H. Seppa, H. Sipola, C. R. Dietlein, and E. N. Grossman, "Passive Euro-American terahertz camera (PEAT-CAM): passive indoor THz imaging at video rates for security applications," in *Proceedings of SPIE*, (Orlando, FL, USA), pp. 654808–654808–7, 2007. 97

[74] A. Luukanen, E. N. Grossman, A. J. Miller, P. Helisto, J. S. Penttila, H. Sipola, and H. Seppa, "An Ultra-Low noise superconducting Antenna-Coupled microbolometer with a Room-Temperature Read-Out," *IEEE Microwave and Wireless Components Letters*, vol. 16, pp. 464–466, Aug. 2006. 97

[75] D. P. Blair and P. H. Sydenham, "Phase sensitive detection as a means to recover signals buried in noise," *Journal of Physics E: Scientific Instruments*, vol. 8, p. 621, 1975. 97, 102

[76] G. Erdi, "Amplifier techniques for combining low noise, precision, and High-Speed performance," *IEEE Journal of Solid-State Circuits*, vol. SC-16, pp. 653–662, Dec. 1981. 97, 100

[77] P. L. Richards, "Bolometers for infrared and millimeter waves," *Journal of Applied Physics*, vol. 76, pp. 1–24, July 1994. 98, 103

[78] J. Schulman and D. Chow, "Sb-heterostructure interband backward diodes," *Electron Device Letters, IEEE*, vol. 21, no. 7, pp. 353–355, 2000. 98, 101, 107, 113

[79] C. Dietlein, J. Chisum, M. Ramirez, A. Luukanen, E. Grossman, and Z. Popović, "Integrated microbolometer antenna characterization from 95-650 GHz," in *IEEE/MTT-S International Microwave Symposium, 2007*, pp. 1165–1168, 2007. 98, 101, 107, 110, 113, 114, 115

[80] H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," *Physical Review*, vol. 83, p. 34, July 1951. 98, 100

[81] P. Helisto and H. Seppa, "Measurement uncertainty in the presence of low-frequency noise," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 2, pp. 453–456, 2001. 98

[82] D. R. White and J. F. Clare, "Noise in measurements obtained by sampling," *Measurement Science and Technology*, vol. 3, no. 1, pp. 1–16, 1992. 99, 101, 102, 104

[83] B. Oliver, "Thermal and quantum noise," *Proceedings of the IEEE*, vol. 53, no. 5, pp. 436–454, 1965. 99

[84] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical Review*, vol. 32, p. 97, July 1928. 99

[85] P. Kittel, "Comment on the equivalent noise bandwidth approximation," *Review of Scientific Instruments*, vol. 48, no. 9, p. 1214, 1977. 99

[86] A. C. Young, J. D. Zimmerman, E. R. Brown, and A. C. Gossard, "1/f noise in all-epitaxial metal-semiconductor diodes," *Applied Physics Letters*, vol. 88, no. 7, p. 073518, 2006. 100

[87] H. Kazemi, C. Nguyen, B. Brar, G. Rebeiz, G. Nagy, L. Tran, A. Young, and E. Brown, "Low cost modular integrated horn antenna array using heterojunction barrier diode detectors," in *IEEE MTT-S International Microwave Symposium Digest, 2008*, pp. 297–300, 2008. 101, 110, 113, 114, 115

[88] J. D. Zimmerman, E. R. Brown, and A. C. Gossard, "Tunable all epitaxial semimetal-semiconductor schottky diode system: ErAs on InAlGaAs," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 23, no. 5, p. 1929, 2005. 101, 109

[89] R. H. Dicke, "The measurement of thermal radiation at microwave frequencies," *Review of Scientific Instruments*, vol. 17, no. 7, p. 268, 1946. 103

[90] H. Kazemi, G. Nagy, L. Tran, E. Grossman, E. Brown, A. Gossard, G. Boreman, B. Lail, A. Young, and J. Zimmerman, "Ultra sensitive ErAs/InAlGaAs direct detectors for millimeter wave and THz imaging applications," in *IEEE/MTT-S International Microwave Symposium, 2007*, pp. 1367–1370, 2007. 107, 114, 115

[91] H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers*, vol. 15 of *MIT Radiation Laboratory Series*. New York, USA: McGraw-Hill Book Company, Inc., 1948. 109

[92] D. Wikner and E. Grossman, "Demonstration of a passive, low-noise, millimeter-wave detector array for imaging," in *Proceedings of SPIE*, (Orlando, FL, USA), pp. 730909–730909–6, 2009. 110, 111, 113, 114

[93] E. R. Brown, H. Kazemi, A. C. Young, J. D. Zimmerman, T. L. J. Wilkinson, J. E. Bjarnason, J. B. Hacker, and A. C. Gossard, "High-sensitivity, quasi-optically-coupled semimetal-semiconductor detectors at 104 GHz," in *Proceedings of SPIE*, (Orlando, FL, USA), pp. 62120S–62120S–7, 2006. 113, 114, 115

[94] R. Meyers, P. Fay, J. Schulman, S. Thomas, D. Chow, J. Zinck, Y. Boegeman, and P. Deelman, "Bias and temperature dependence of sb-based heterostructure millimeter-wave detectors with improved sensitivity," *Electron Device Letters, IEEE*, vol. 25, no. 1, pp. 4–6, 2004. 113, 114, 115